Editors Jim Graham, Neil Thacker and Tim Cootes are employees of the University of Manchester.

ISBN 1 901727 31 6

Steering Commitee

Referees

Preface

This is the 10th conference on Medical Image Understanding and Analysis. Since 1997 MIUA has served as a forum for presentation and discussion of scientific research in the area of medical imaging and the analysis of medical images among researchers coming from a range of disciplines – engineering, physics, computer science, clinical practice and fundamental bioscience. While researchers in each of these communities have their own outlets for publishing their work, MIUA provides a focus for their common interests. MIUA's role is communication among researchers in the U. However, it has had an increasing participation from overseas, and that trend continues this year with contributions from the Netherlands, Poland, USA and Canada.

Since the remit of MIUA is communication amongst members of the UK community, the principal criteria are scientific soundness and interest, rather than originality. Work can, and has, been accepted for MIUA that has been published elsewhere, provided it merits presentation to the UK community.

As organisers we call on assistance from others in the numerous duties involved in organising a conference. Our thanks to all of them, but in particular we can single out Mike Rogers who set up the web site and on-line administration, Simon Jackson who provided CAWS cover when Mike left and Christie Finegan for administrative help. Of course we are grateful to the referees who (under a bit of time pressure) provided the opinions for selection of the conference papers. Refereeing is a difficult and thankless, but necessary task, without which no conference could function. While we cannot guarantee that, in setting acceptance thresholds on the referees assessments we have not rejected scientifically valid contributions, we feel confident that what has been accepted will interest and stimulate you.

Jim Graham, Tim Cootes, Neil Thacker, Sue Astley.

Table of Contents for Volumes 1 and 2

Podium Session 3: Bones

Invited Presentation

Podium Session 4: Methods

Podium Session 5 : Models

Podium Session 6: Organs

Poster Session 2

Large-Scale Image Analysis for High-Content Screens in Cell Biology

Roland Eils

Division of Theoretical Bioinformatics, German Cancer Research Centre (DKFZ) and Institute of Pharmacy and Molecular Biotechnology (IPMB), University of Heidelberg, Germany

Microscopy of cells since its early days in the mid-seventeenth century has changed dramatically. Methods of image analysis have changed from measurements of hand drawings and still photographs to using computational methods to (semi-) automatically quantify numbers of objects, distances, concentrations and motion velocities of cells and sub-cellular structures. Techniques like contrast enhancement with Differential Interference Contrast (DIC), phase contrast, confocal or multiphoton microscopy as well as probe detection by fluorescent markers are today commonly employed in studying cellular structures and dynamics. Since the discovery of the green fluorescent protein (GFP) as a universal marker for proteins in the living cell, live cell imaging technologies using GFP and its spectral mutants have revolutionized the study of cellular dynamics. Imaging of dynamic processes in the living cell involves microscopic techniques with high-resolution both in space and time. Such imaging technologies generate a wealth of data that requires multi-dimensional image analysis, visualization and quantitative image analysis as a prerequisite to turn qualitative and descriptive data into quantitative values. Such quantitative data is the basis for mathematical models of protein kinetics and quantitative biochemical and signaling networks, and opens the way towards a quantitative rather than a purely descriptive view on cell biology.

After the completion of the human genome sequencing project the task of functional genomics is to discover protein function genome-wide. Currently, RNA interference (RNAi) is the method of choice to study loss of function phenotypes in human cells by specifically suppressing the expression of virtually any desired protein coding gene. Using automated microscopy a genome-wide screen for approx. 25.000 genes can be conducted in a few days only (Neumann et al. 2006). These screens typically produce 2-3 TBytes of compressed image data that clearly requires automated tools for image analysis and classification (Conrad et al. 2004). In my presentation, I will demonstrate methods for fully automated image analysis and interpretation of large-scale data produced in functional gene knockdown experiments on live cell arrays (Harder et al. 2006). In particular, I will introduce a computational scheme to process 3D multi-cell time-lapse images. This approach is used to automatically segment, track, and classify cell nuclei into different phases of the cell cycle. Our scheme proves a high accuracy, suggesting a promising future for automating the evaluation of high-throughput experiments in molecular cell biology.

References

1. C. Conrad, H. Erfle, P. Warnat, N. Daigle, T. Lörch, J. Ellenberg, R. Pepperkok, and R. Eils (2004) Automatic identification of subcellular phenotypes on human cell arrays,. Genome Research, vol. 14, pp. 1130-1136, 2004.
2. N. Harder, B. Neumann, M. Held, U. Liebel, H. Erfle, J. Ellenberg, R. Eils, and K. Rohr (2006) Automated recognition of mitotic patterns in fluorescence microscopy images of human cells. in Proc. IEEE Int. Symp. Biomedical Imaging (ISBI'2006), Arlington, USA, 2006, pp.1016-1019.
3. B. Neumann, M. Held, U. Liebel, H. Erfle, P. Rogers, R. Pepperkok and J. Ellenberg (2006) High-throughput RNAi screening by time-lapse imaging of live human cells. Nature Biotech, in press.

Prone-Supine Breast MR Image Registration for Image-Guided Surgery

T.J. Carter, C. Tanner, W.R. Crum, D.J. Hawkes

Centre for Medical Image Computing (CMIC), University College London, London, WC1E 6BT, UK

{t.carter, c.tanner, b.crum, d.hawkes}@cs.ucl.ac.uk

Our aim is to use contrast enhanced MR images for image-guided breast surgery. Unfortunately, significant deformation of the breast occurs between the prone position, in which diagnostic MR images are acquired, and the supine position, in which surgery is performed. This deformation substantially reduces the proximity of corresponding image features so intensity-based image registration algorithms perform poorly. We present a technique to register prone and supine breast MR images based on the initialization of an intensity-based deformation by a finite element model. The finite element model, constructed from the supine image, is deformed to match the prone image by simulating the effects of gravity and by matching the skin surface to the prone image. The supine image is deformed according to the model displacements, and the resulting image is registered to the prone image using a fluid registration algorithm. This technique can recover the displacement of landmarks with an RMS target registration error of 3.0mm, compared with an error of 10mm for a standard non-rigid registration technique.

1 Introduction

The most common treatment for breast cancer is breast-conserving surgery, in which the surgeon removes the lesion and a small margin of healthy tissue surrounding it, but preserves the majority of the breast tissue. A significant proportion of these operations require re-excision, predominately due to involved margins being found when the excised tissue is examined histologically. The position and extents of a breast lesion are indicated as a region of enhancement in dynamic contrast enhanced magnetic resonance images, which are acquired with the patient in a prone position. However the breast is a very soft, deformable tissue and so significant deformation of the breast occurs between being imaged prone and the surgery which is performed in a supine position.

Successful results have been reported for using intensity-based non-rigid registration techniques to register prone breast MR images together [1][2]. However these techniques are less successful in situations where there is not a good overlap of consistent image features, which will occur, for example, when there has been significant deformation between images. We have found existing intensity based methods are unable to capture the large deformation which occurs between prone and supine breast images. The challenge of prone-supine breast registration has been addressed by Behrenbruch et al. [3] who use a B-spline tensor mesh to perform a non-rigid registration driven by the skin surface. However they demonstrate accuracy to within the nearest quadrant only.

In this paper we propose a novel technique for performing registration of MR images in situations where significant deformations have occurred. Fluid registration techniques [4] have previously proved successful at recovering large displacements but can be susceptible to local minima. Therefore in this work we use a finite element model to initialise a fluid registration of the prone to supine case. We illustrate our framework with an example dataset, where we perform registration between images of the breast acquired of a patient in prone and supine positions.

2 Method

2.1 Overview

A finite element model, constructed from the supine image, was deformed to approximate the deformation of the breast due to gravity between the supine and prone positions. The gravity-deformed model was rigidly aligned with the prone surface and then deformed by imposing displacements on its surface to match the breast surface in the prone image. The finite element displacements were used to construct an estimate of the prone image from the supine image. Finally an intensity-based non-rigid registration algorithm was used to align this estimate with the prone image, establishing point-by-point correspondence between the prone and supine image pair.

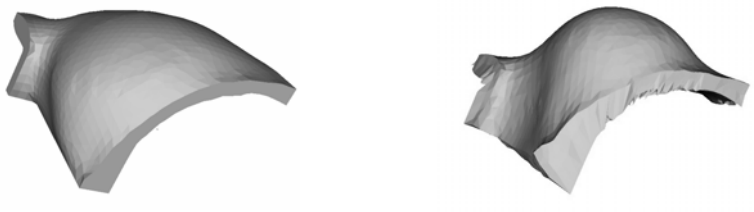
2.2 Image Data

Prone and supine images were acquired of a symptomatic breast patient on a Philips 1.5 T MR system. Nine MR-visible fiducial markers (MM3002, IZI Medical Products Corp, Baltimore) were affixed to the breast. The prone images were acquired using a dedicated breast coil, according to the standard clinical protocol. A 3D gradient echo

sequence (TR/TE/flip = 20ms/5.2ms/45˚) was used to acquire coronal images (0.7mm x 0.7mm x 2.2mm). The supine images were acquired axially (to reduce breathing motion artifacts) and the Q-body coil was used so that there were no additional mechanical forces on the breast. A 3D gradient echo sequence (TR/TE/flip = 14ms/4.1ms/25˚) was used to acquire the axial images (0.7mm x 0.7mm x 2.5mm).

2.3 Model

We constructed our initial model in the supine position (Figure 1a). This position was chosen as we believe the deformation from supine to prone will be more predictable, since the breast is moving to a position in which it hangs freely. The 3D MR image of the supine breast was segmented into fatty tissue, fibroglandular tissue and background using the Analyze software package (http://www.mayo.edu/bir). The image was isotropically sub-sampled and blurred to increase the smoothness and the element size such that the limit of available elements was not exceeded during the meshing stage. Surface meshes were extracted for the skin and chest wall using the Visualisation ToolKit (http://public.kitwarecom/VTK). These surfaces were connected by planes along their edges. This volume was meshed into ten-noded tetrahedral elements using ANSYS (http://www.ansys.com). Each element was labelled according to the tissue segmentation. A layer of 1mm thick six-noded triangular prisms was added to the model to represent skin. The fatty, fibro-glandular and skin tissues were modelled as isotropic, linear materials with Young's Moduli of 0.8kPa, 1.2kPa and 10kPa respectively [5,6]. All materials were assigned Poisson's ratios of 0.495 for nearly incompressible tissue. The density of all materials was modelled as 1000kg/m^3. In the model the node closest to each fiducial marker on the supine MR image was identified as a fiducial node.



(a) (b)

Fig. 1. (a) Surface rendering of model built from supine image and (b) model after the supine to prone displacement due to gravity has been simulated

2.4 Gravity

Initial stresses, and strains, should be included in the model built from the supine image, since gravity was acting on the breast when it was imaged. We approximated the initial stress in the vertical direction as increasing linearly with depth (which is determined from the MR image) due to the weight of tissue above it, whilst all other stresses were considered to be zero. Gravity was then applied to the model to deform it to a prone position. The model was constrained by imposing zero-displacements on nodes along the chest wall and was solved using an infinitesimal strain formulation. From this gravity-deformed supine model we created a prone model, in which all initial stresses were set to zero and on which no gravity was acting (Figure 1b). This greatly simplifies the process of imposing suitable boundary conditions to constrain the prone model during the next step, in which the surfaces are aligned.

2.5 Surface alignment

The triangulated surface of the prone image was extracted from the MR image using the same techniques as for the supine image. The vertex normal (the mean of the normals of the attached areas) was calculated for each vertex. The model in the prone position was rigidly registered with the prone surface by aligning the corresponding fiducial markers. Displacement loads were then applied to each of the fiducial nodes to position them in the same location as the corresponding fiducial marker in the prone image.

The model was aligned with the target surface by imposing displacement loads on its surface using an iterative method. For each iteration the closest point on the prone surface to each node on the skin surface of the finite element model was calculated. The local coordinate system at each of these nodes was rotated such that the z-axis aligned with the vertex normal. A displacement load was then applied along this axis, but no constraints were applied in the other directions so that the point could slide along the surface tangent. The distance that the node was displaced along the z-axis was proportional to the length of the vector from the model node to the prone surface vertex, resolved into the direction of the vertex normal (Figure 2). The nodes along the chest wall were constrained

to act together as a rigid body. The model was solved at each iteration using ANSYS, and the node positions and model stresses were updated.

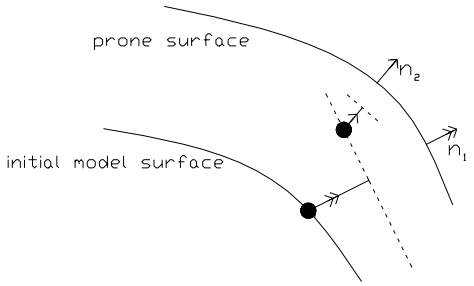


Fig. 2. At each iteration, a model surface node(●), was constrained to move a set distance in the vertex normal direction n_i of the nearest prone surface vertex, but could slide parallel to the surface tangent (marked with a dotted line).

2.6 Fluid registration

The calculated finite element displacements were applied to the supine image in order to create an image which was approximately aligned with the prone image. These two images were registered using a fluid algorithm [1], in which the transformation between two images is modelled as the flow of a compressible viscous fluid [4]. This framework produces diffeomorphic transformations that can accommodate large deformations. We used normalised intensity cross correlation as the image similarity measure to drive the registration with viscosities, μ and λ, set to 1 and 0 respectively as in previous work. The regridding threshold (the fractional change in voxel volume below which the current transformation is applied and the algorithm restarted) was set to 0.5. A stopping condition based on the change in similarity measure at each iteration was used to terminate the registration when there was no further improvement in image similarity, with the maximum possible number of iterations set to 200. The algorithm was run at half image resolution level for speed and robustness.

3 Results

The significant deformation between prone and supine images can be clearly seen in Figures 3a and 3b. The majority of the deformation visually appears to be recovered by our proposed prone-supine registration algorithm (Figure 3c). To quantify the accuracy of our prone-supine registration technique five landmarks were manually identified in the prone and supine images within the volume of the breast in order to calculate a target registration error (TRE). The root mean square (RMS) TRE was 3.0mm and the maximum TRE was 3.7 mm. The TRE was calculated for several other possible registration techniques: performing a rigid registration only; performing a rigid registration followed by a finite element deformation as described in sections 2.4 and 2.5; and performing a rigid registration followed by a fluid registration only. The proposed technique had the lowest TRE, as can be seen in Figure 4.

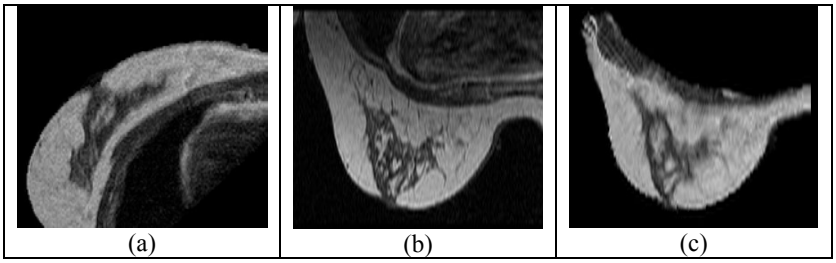


Fig. 3. 2D example slice of (a) supine image, (b) prone image (c)supine image deformed by finite element model followed by fluid registration with prone image

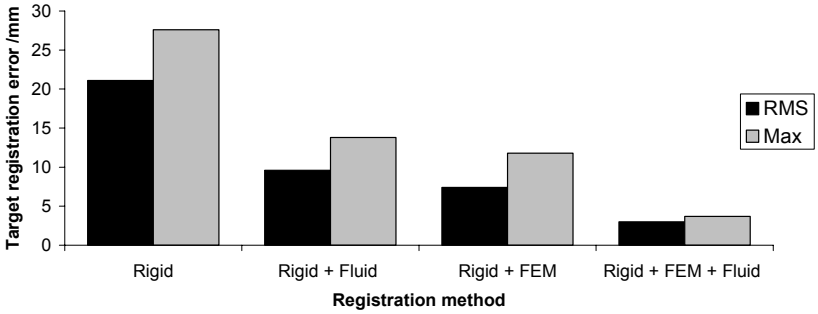


Fig. 4. Comparison of TREs for prone-supine registration using the proposed method (Rigid+FEM+Fluid) with rigid registration only (Rigid); rigid registration followed by fluid registration only (Rigid + Fluid); and rigid registration followed by only the finite element alignment steps described above (Rigid + FEM).

4 Discussion

Intensity based registration generally performs poorly in situations where significant deformation has occurred. A number of strategies have been developed to make it more robust in other situations where corresponding features are not initially close to one another including multi-resolution approaches and the incorporation of information such as corresponding landmarks or regions. We have presented here a new technique to perform registration of prone and supine MR images of the breast based on a finite element model-based initial estimate of the displacement, followed by an intensity-based non-rigid registration. We have demonstrated our technique by registering a prone breast MRI with a supine image.

On the example case we present here, we have found our technique to be capable of recovering landmark displacements with an accuracy of 3.0mm RMS. It should be possible to improve the accuracy of the model by considering a finite strain formulation and non-linear material properties. In future work we will investigate whether such changes, based on the limited image and elasticity information available, would improve the overall registration result, or whether the effect is small enough that it is being recovered in the fluid registration step. The initial stress in the model is similarly only crudely estimated currently, and we intend to implement a more rigorous technique to estimate it.

The reported accuracy is likely to be sufficient for our target application of breast surgery, where 10mm margins around a cancer are frequently allowed, and it is substantially more accurate than has been previously reported for the registration of prone to supine breast images. After further validation work, we intend to use this registration technique as part of the image-guide breast surgery system which we are developing.

Acknowledgements

We would like to thank the surgeons, radiologists and radiographers at Guy's Hospital for their assistance in this research. Carter, Tanner and Crum acknowledge the support of the Medical Images and Signals IRC (EPSRC GR/N14248/01 and UK Medical Research Council Grant No. D2025/31).

References

1. Crum W.R. et al.: Multiresolution Anisotropic Fluid Registration: Evaluation in Magnetic Resonance Breast Imaging, Physics in Medicine and Biology 50, pp 5153-5174, 2005.
2. Rueckert D et al.: Non-rigid Registration using Free-Form Deformations: Application to Breast MR Images. IEEE Transactions on Medical Imaging, 18(8):712-721, 1999.
3. Behrenbruch C.P. et al.: Prone-Supine Breast MRI Registration for Surgical Visualisation. Medical Image Understanding and Analysis, 2001.
4. Christensen, G.E. et al.: Deformable templates using large deformation kinematics", IEEE Transactions on Image Processing, 5 (10): 1435-1447, 1996
5. Samani A. et al.:. Biomechanical 3-D Finite Element Modeling of the Human Breast Using MRI Data IEEE Transactions on Medical Imaging, 20:271-279
6. Sarvazyan A. et al.:. Elastic Imaging as a New Modality of Medical Imaging for Cancer Detection. In Proc Int Workshop on Interaction of Ultrasound with Biological Media, Valenciennes, France, pages 69-81, 1994

Classification of Meningiomas using Discriminant Wavelet Packets and Learning Vector Quantization

Hammad A. Qureshi[a], Nasir M. Rajpoot[a], Khalid Masood[a], Volkmar Hans[b*]

[a]Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK
[b]Institute of Neuropathogy, Evangelisches Krankenhaus, Bielefeld, Germany

Abstract. This paper presents a novel texture-based algorithm for detecting certain kinds of meningiomas in images of neurosurgical resections. The algorithm employs Discriminant Wavelet Packet Transform (DWPT) and Learning Vector Quantization (LVQ). The adaptive DWPT of a test image is computed by maximizing the discrimination power of subbands during the basis selection process for the training images. The discrimination power of a subband is computed with the help of Hellinger distance between pseudo probability density functions of the subband. Statistical features are then obtained for each of the top few most discriminant subbands and finally LVQ is trained and subsequently used for detection. The proposed methodology produces promising results for the multiple class meningioma classification problem with near-perfect results for detection of certain kind of meningiomas.

1 Introduction

Meningiomas are tumours of the brain and the nervous system. They arise from the cells of the meningeal covering of the brain and the spinal cord. Meningiomas account for 20% of all brain tumours and exist in the three membranes covering the brain and the spinal cord. The majority of meningiomas are benign but they can still be fatal because as they grow they push against other brain tissue. As per the World Health Organization (WHO) recommendations the meningiomas have been classified into six different types: Meningiothelial, Fibroblastic, Transitional, Psammomatous, Angiomatous and Chordo. There has been a lot of concern lately pertaining brain tumors, as according to one estimate, as many as 6 persons in 100,000 people may have some from of meningioma. The data acquired from meningioma patients at a certain instant may be huge in volume which would cause the neurosurgeon and neuropathologist to spend valuable time studying them. Therefore, its extremely important that a computerized technique is developed that would facilitate the handling of large amount of data and analysis of complex images in a reasonable time frame. This would lead to the reduction of costs and would mean access to timely treatment for more patients. Moreover, for a minority of cases around 10 to 20 percent occurrences of the said disease, the prognosis has been found to be difficult. A computerized technique may be able to help the neurosurgeon and neuropathologist involved in such cases in reaching a conclusion about the meningioma and would act as a decision support system. Hence a medical practitioners job is made easier and process of diagnosis is quicker.

Meningiothelial	A synctium is formed of the cells, lobulated
Fibroblastic	Spindle shaped cells and a matrix is abundantly found in the collagen
Transitional	Whorls are found with a few psammoma bodies, fibroblastic features present
Psammomatous	Appear a bit transitional with psammoma bodies dispersed

Table 1. Description of the four WHO grade I Meningioma subtypes analyzed in this study

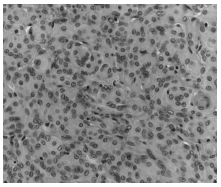
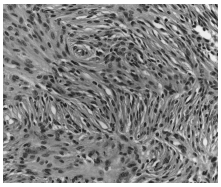
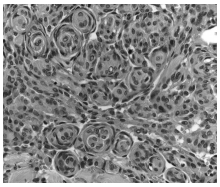
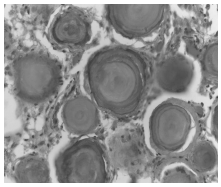
a.  b.  c.  d. 

Figure 1. Meningioma Images for each subtype a. Meningiothelial, b. Fibroblastic, c. Transitional, d. Psammomatoes

In this study, data for only four out of six meningiomas (described in Table 1) was available for analysis. The data acquisition process involved the analysis of the resection slides on a Zeiss Axioskop 2 plus microscope with a Zeiss

*hammad@dcs.warwick.ac.uk, nasir@dcs.warwick.ac.uk, khalid@dcs.warwick.ac.uk, volkmar.hans@evkb.de

Achroplan 40x/0,65 lens to obtain 1300×1030 pixels, 24 bit, true color RGB pictures. Some of the meningioma images of the four subtypes considered in this study are shown in Figure 1. These images are converted from their RGB values to 8-level grayscale images for analysis.

Recently some work has gone into the detection and classification of brain tumors. One such effort in the domain of meningiomas was carried out by Lessman et al. [1] who studied the problem of content-based visualization of meningioma images to aid in characterization of the database contents. In this paper, we combine the feature extraction capability of the Discrete Wavelet Packet Transform (DWPT) based upon the selection of the subbands that are most useful for texture classification with classification ability of neural networks. The ability of wavelet packet transform in devising a feature selection algorithm for texture classification was demonstrated in [2, 3]. Our work is motivated by the basis selection paradigm presented in [4] and texture classification using discriminant wavelet packet subbands [5]. Statistical parameters [6] of the selected subbands are used as features for training and detection purposes. Learning Vector Quantization (LVQ) [7] is used as a classifier to discriminate between different kinds of meningiomas. LVQ has been used in the literature to discriminate between images based upon texture [8] and to classify pancreatic tissues based on their texture [9]. The following section describes the methodology of the proposed algorithm in detail. The results are presented at the end with a discussion and future directions.

2 Method

Pattern classification and detection is usually composed of three stages. The first is the pattern feature extraction, then training of the neural network on a subset of the feature images and finally the testing of all images to determine, which class they belong to. In this paper we present a technique which is based upon the above paradigm and the methodology is subdivided into three stages as shown in Figure 2.

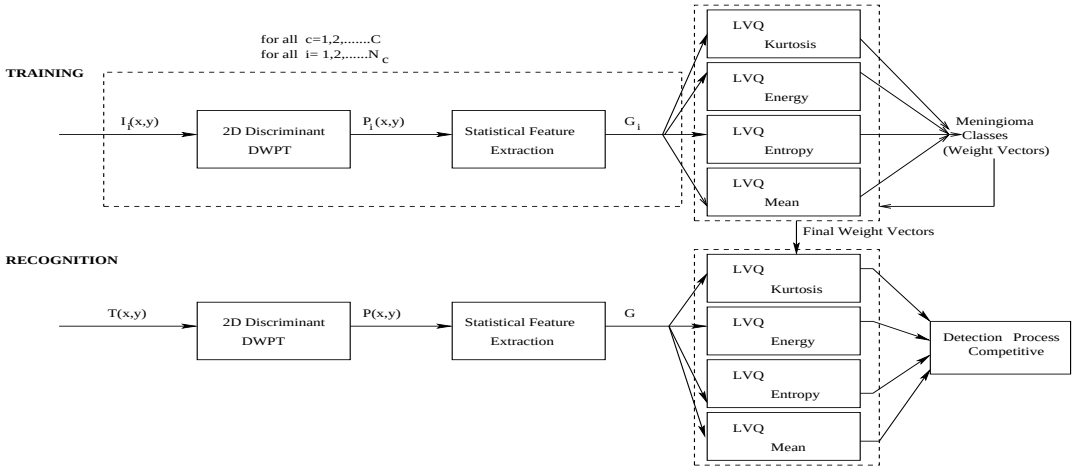


Figure 2. Block Diagram of the Methodology (DWPT=Discrete Wavelete Packet Transform, LVQ=Learning Vector Quantization)

2.1 2D Discriminant Wavelet Packet Basis Selection

The first stage of our technique is the selection of the best basis i.e. the best description of the images in terms of wavelet packet subbands. This is important because we wish to obtain the representation of the image which would aide us in localizing the spatial frequencies and comparing the most differentiating aspects of the images. The process is divided into two steps:

2.1.1 Adaptive Wavelet Packet Transform

The wavelet packet transform is a mechanism by which the high frequency and low frequency components of a signal are decomposed into separate entities. The wavelet transform is computed by applying a certain highpass and a lowpass filter upon the input signal to acquire the high frequency and low frequency subbands. The wavelet packet transform is performed iteratively on every subband to a certain level to obtain the high and low frequency coefficients at each stage. Without loss of generality, the transform for a 1D discrete signal $\mathbf{x} = \{x_i\}, i = 0, 1, ..., N-1$ may be computed as

follows:

$$S_0^0(l) = x_l \quad l = 0, ..., N-1$$

$$S_j^{2p}(l) = \sum_k g_{k-2l} w_{j-1}^n(k); S_j^{2p+1}(l) = \sum_k h_{k-2l} w_{j-1}^n(k) \quad l = 0, ..., 2^{-j}N - 1$$

where g_n and h_n are lowpass and highpass filters respectively, $j = 1, 2, ..., J; J = log_2 N$, $S_j^p(l)$ is the transform coefficient corresponding to the wavelet packet function having relative support size 2^j, frequency $p2^j$ and is located at $l2^j$. Hence, j, p and l are regarded as the scale, frequency and position indices of the wavelet packet function. The transform is invertible if appropriate dual filters $\tilde{g}_n$ and $\tilde{h}_n$ are applied on the synthesis side.

The wavelet packet decompositions are maintained in a quadtree structure, with the parent being the original subband or image and the children being the wavelet decompositions of the parent. First the image is decomposed into its respective subbands and then each subband is decomposed further until a predefined maximum depth of the tree is reached. This results in a combinatorial explosion of possible wavelet packet bases that can be used to completely represent the image. Hence, selection of the best basis is desirable which will adaptively maximize the discrimination power of the decompositions. This is achieved with the help of dynamic programming using an appropriate criterion as described below.

2.1.2 Discriminant Subband Selection

In the selection of subbands, a pseudo probability density function (pdf) is obtained for each subband using the normalized energy for the subband coefficients. Then the discriminating power of each subband is obtained, using the Hellinger distance between different classes, as follows.

$$\mathcal{D}(x_1, x_2) = \sum_i (\sqrt{x_1^i} - \sqrt{x_2^i})^2$$

where x_1^i and x_2^i are the normalized energy of the ith subband coefficient for two training images of different classes. This distance is used as the cost function in the best basis selection process. Once we have all the subbands that are more discriminating and a discrete wavelet packet transform of the image is obtained, a ranking of all the 91 subbands (as shown in Figure 3) according to the discriminating power is obtained. The second stage of our technique is as followed.

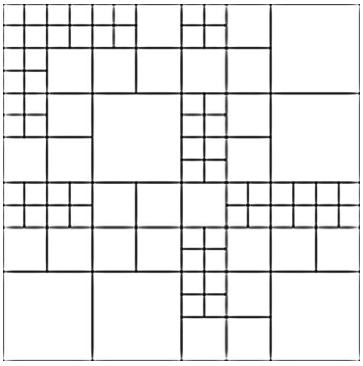


Figure 3. Fixed 4-Level Adaptive Wavelet Packet Geometry of a Meningioma Image for our analysis

2.2 Feature Extraction

As stated above one of the most important aspect of any pattern detection technique is extracting inherent features in an image. Consequently any texture classification technique would only be as good as the features extracted from the pattern being analyzed. The process of feature extraction is compounded in difficulty by the dimensionality of the feature space. The complexity of features selection is reduced to a great extent by employing an adaptive wavelet packet transform and using it to extract the most discriminant subbands from the complete transform. The following statistical parameters (features) were found to be useful for our application: *Kurtosis*: $\mu_4 = \sigma^{-4} \Sigma_{i=0}^{G-1} (i-\mu)^4 p(i) - 3$, *Energy*: $E = \sum_{i=0}^{G-1} [p(i)]^2$, *Entropy*: $H = -\sum_{i=0}^{G-1} p(i) log_2[p(i)]$ where $p(i) = h(i)$ / N $\quad i = 0,1,\ldots,G\text{-}1$, $\mu = \sum_{i=0}^{G-1} ip(i)$, $\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 p(i)$, G is the number of pixel intensity levels and $h(i)$ is the histogram count of pixel intensity i in

the subband. These parameters are computed for the probability density function of each subband and are subsequently used to train the neural network.

2.3 Neural Network and Pattern Classification

LVQ is a supervised network, which approximates the decision surfaces of the theoretical Bayes classifier. The training and usage of LVQ is described below and this is the final stage of the method.

2.3.1 Training of Neural Network-Learning Vector Quantization

In the training phase the statistical features (mean, kurtosis, energy, entropy) of the subbands are used as input vectors. Multiple LVQs are trained and feature vectors for selected subbands are used as training data. Each output of a LVQ represents a particular class or category. At the start of the training process, a weight vector is selected which represents the reference vector for the class that the unit represents. During training, the output units are modified through supervised training. As per the mechanism, the reference vectors are represented by the initial weights $W_i = w_{i1}, w_{i2}, w_{i3}, ..., w_{ij}$ where W_i represents the weight vector for a certain class and j denotes the number of elements in each weight vector. In our case they represent the number of subbands selected for our analysis at each stage. These weights are adjusted at each stage of training called the epoch by the rule $W_i = W_i + \alpha_t(V_k - W_i)$ if the input vector V_k has been classified correctly or by $W_i = W_i - \alpha_t(V_k - W_i)$ if the V_k has been classified incorrectly. The parameter α_t is defined as the learning rate which is adjusted at each stage, a typical value for α_t is 0.1 and it is reduced by 0.001 at each stage of training. The training stops when α_t reaches zero or a certain criteria for training is met such as all the input vectors are classified correctly. After the training, LVQ net classifies each input vector according to the weight vector (reference vector) to which it is closest. The proximity of a certain vector to a certain class is determined by its euclidean distance i.e. how close the reference vector is to the vector to be classified $d = |W_i - V_k|$, where d is a measure of proximity.

At the training stage, we selected 5 samples of each kind of meningiomas and employed them for our training purposes. Hence, after training we acquired a set of weight vectors which represented each class in the training samples separately. There were in all 84 sample images, 20 were used for training and the other 64 were used as test images. This effort was a first experiment with heuristically derived training set of 20 meningioma images. A more detailed analysis with cross validated training data would be provided soon.

2.3.2 Detection and Classification of Meningiomas

As shown in Figure 2, the detection process involves a process of competition. Four neural networks are trained based upon different statistical features (mean, kurtosis, energy and entropy). Different length of feature vectors are used for each LVQ based upon the efficiency of the results (for mean 56 subbands were used, for kurtosis 82, energy 68 and for entropy 58), the relevant features are computed for each subband. The block diagram 2 shows three neural networks because the two different LVQs for entropy and mean are shown as combined, in actual process they are separate. Then for each test image, the closest class is computed by each LVQ. If more LVQs would detect the current image as a certain class, it would finally be selected as the class to which the meningioma belongs. The reason for employing different statistical features is that each one of them was found to be better able to identify a certain meningioma. For instance, Kurtosis of the subbands was better able to identify meningothelial and other features were useful in detecting other meningiomas. A LVQ trained on kurtosis would be able to detect other meningiomas as well but with less efficacy.

3 Results and Discussion

The experimental setup consisted of employing the Daubechies 4 tap filter for obtaining the wavelet packet transform upto four levels. The data for other kinds of meningiomas, namely, Angiomatous and Chordo were not available to us at this time. The process on the whole involved: (a) adaptive wavelet packet transform of meningioma images, (b) statistical features extraction of the wavelet packet decomposition, (c) train multiple LVQs (neural networks) with extracted features (d) perform recognition using LVQ. The results of our study are depicted in Table 2.

The proposed algorithm has shown some promising results. Despite being simple and fast, the LVQ classifier has achieved high detection rate for the meningothelial and psammomatous meningiomas. This is encouraging as meningothelial is the one of the most common form of the tumor while psammomatous is the most developed because in it the cancerous cells surround themselves with protein and form psammoma bodies. Our feature selection mechanism fa-

Meningioma	Correct Detections	Total Test Images	Classification Rate
Meningiothelial	18	19	95%
Psammomatous	15	15	100%
Fibroblastisch/Transitional	15	30	50%
Overall	48	64	75%

Table 2. Meningiomas Classification results.

vored the detection of these particular kind of meningiomas. However, more sophisticated statistical features and advanced feature selection may be desirable to do as well on all kinds of meningiomas.

In certain instances fibroblastic tumor was found to resemble meningothelial to a great extent. Therefore it has been difficult to detect these as exclusively as desired (as depicted by the results). Meningioma detection process was compounded in difficulty by the fact that certain images classified as one kind contained textures belonging to more than one kinds of tumor. This was most significantly observed for the fibroblastic and transitional meningioma samples. Fibroblastic in certain images resembled meningothelial to a great extent and the transitional as stated earlier does contain fibroblastic and meningiothelial features by definition. Therefore, the one meningiothelial image classified wrongly is classed as transitional and the transitional and fibroblastic classified wrongly are classified as mengiothelial.

4 Conclusions

In this paper, we have presented a novel texture based algorithm for detection and classification of meningioma images. The algorithm was based on selecting the most discriminant wavelet packet subbands and using certain statistical features of the subbands for training the neural network. The experimental results demonstrated the effectiveness of the algorithm while exhibiting its limitations in differentiating between fibroblastic and transitional type of meningiomas. Future work may look at the effect of different wavelet filters on the classification performance. The selection of discriminant wavelet packet basis for a particular pair of meningioma classes may be useful in efficiently classifying images of different meningiomas with neural networks. Different kinds of neural networks and other classifiers may be investigated for better performance.

Acknowledgements

The authors would like to thank Prof Tim Nattkemper of the University of Bielefeld (Germany) for sharing his insights into the problem.

References

1. B. Lessmann, V. Hans, A. Degenhard et al. “Feature space exploration of pathology images using content-based database visualization.” In *Proceedings SPIE Medical Imaging*. 2006.
2. N. Rajpoot. “Texture classification using discriminant wavelet packet subbands.” In *Proceedings 45th IEEE Midwest Symposium on Circuits and Systems (MWSCAS)*. 2002.
3. A. Bhalerao & N. Rajpoot. “Discriminant feature selection for texture classification.” In *Proceedings 14th British Machine Vision Conference (BMVC)*. 2003.
4. N. Rajpoot, R. Wilson, F. Meyer et al. “A new basis selection paradigm for wavelet packet image coding.” In *Proceedings IEEE International Conference on Image Processing (ICIP)*. 2001.
5. N. Rajpoot. “Local discriminant wavelet packet basis for texture classification.” In *Proceedings SPIE Wavelets X, San Diego, California*. 2003.
6. A. Materka & M. Strzelecki. “Texture analysis methods - a review.” In *COST B11 report, Institute of Electronics, Technical University of Lodz*. 1998.
7. L. Fausett. *Fundamentals of Neural Networks*. Prentice-Hall, 1994.
8. S. Livens, P. Scheunders, G. Van de Wouwer et al. “A texture analysis approach to corrosion image classification.” *Microscopy, Microanalysis, Microstructures* **7(2)**, pp. 143–52, 1996.
9. T. Mattfeldt & H. A. Kestler. “Classification of pancreatic tissue texture on the basis of correlation dimensions and Lyapunov exponents using two types of artificial neural networks.” In *Interdisciplinary approaches in fractal analysis (IAFA)*. 2003.

Mammographic Risk Assessment Based on Anatomical Linear Structures: Do Mammographic Abnormalities Play a Role?

Edward M. Hadley[a], Erika R. E. Denton[b] and Reyer Zwiggelaar[a*]

[a]Department of Computer Science, University of Wales, Aberystwyth, UK,
[b]Department of Radiology, Norfolk and Norwich University Hospital, UK

Abstract. Mammographic risk assessment is concerned with the probability of a woman developing breast cancer. Recently, it has been suggested that the density of linear structures is related to risk. For 321 images from the MIAS database, a measure of line strength was obtained for each pixel using the Line Operator method. The proportion of pixels with line strength above a threshold level was calculated for each image and the results categorised by Tabar pattern and Boyd SCC class, and the differences between results for the set of all images and the subset of 206 normal images were investigated. The results indicated a significant difference between Boyd classes 1–3 (low risk) and classes 4–6 (high risk), and between most Tabar patterns. No significant differences were found between the results for the set of all images and the set of normal images.

1 Introduction

Mammographic risk assessment is concerned with estimating the probability of women developing breast cancer. Risk assessment is a rapidly developing area of research and aims to improve the likelihood of the early detection of breast cancer. Breast density is an important indicator of mammographic risk [1] and the best predictor of mammographic sensitivity [2]. However, more recently, it has been suggested that the distribution of linear structures is also correlated with mammographic risk [3–5]. So far it is not entirely clear if it is just the density of linear structures (either by percentage area or volume) or if the distribution of the linear structures plays a role as well.

Tabar et al. have proposed a mammographic risk assessment model based on four structural components, where the relative proportions of each component is linked to the risk of developing breast cancer [3–5]. One of the four structural components is linear density. The main purpose of this work is to investigate if automatic methods can be used to correlate the density of linear structures to mammographic risk classification metrics.

Two classification models are used: Tabar patterns [5] and Boyd SCC classes [6]. Tabar's classification consists of five *patterns*, where patterns I–III represent a low risk of developing breast cancer, and patterns IV–V indicate a higher risk. Screening tests have shown that cancer prevalence in women with patterns IV-V is approximately twice that in women with patterns I-III [5]. The Boyd SCC model consists of a scale of six classes where class 1 indicates the lowest risk and class 6 indicates the highest risk. Since the Boyd model is based on density, correlation with the Boyd model can be seen as an indirect correlation with breast density. BIRADS classes [7] were also used in a more detailed anlysis of the initial results [8].

2 Method

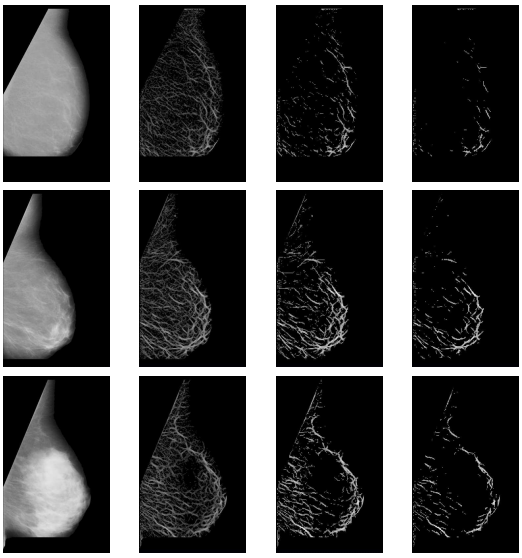
Three hundred and twenty-one mammographic images from the Mammographic Image Analysis Society (MIAS) database were classified according to Tabar patterns [5] and Boyd SCC classes [6] by an expert radiologist. Example images of low, moderate and high risk mammograms are shown in Fig. 1 (a).

The images were processed using Dixon and Taylor's line operator method [9, 10] (see Sect. 2.1), producing a measurement of line strength at each pixel. The multi-scale line operator has been shown to be more effective at detecting linear structures in mammographic images than other methods [10].

Figure 1 (b) shows examples of low, moderate and high risk mammograms following processing with the line operator.

The relative proportion of pixels with line strength values above a range of thresholds was calculated for each image. Figure 1 (c, d) shows examples of the resultant images after thresholding. Subsequently, the results were analysed for differences between images of each Tabar pattern and Boyd SCC class.

*Edward Hadley: emh05@aber.ac.uk, Reyer Zwiggelaar: rrz@aber.ac.uk.



(a) (b) (c) (d)

Figure 1. The top row shows a mammogram of Boyd SCC class 1/Tabar pattern II (low risk), the middle row shows a mammogram of Boyd SCC class 3/Tabar pattern III (moderate risk) and the bottom row shows a mammogram of Boyd SCC class 6/Tabar pattern IV (high risk). The images in column (a) show the original mammograms, column (b) shows the results after processing with the line operator, and columns (c) and (d) show the results after thresholding at 4/204 and 6/204 respectively. The lines in (b), (c) and (d) have been enhanced for viewing.

Following this analysis, the subset of normal (non-cancerous) images was selected and the results for this set compared with the results for the set of all images. The full set of 321 mammographic images contained 206 normal and 115 abnormal images. Graphs were produced showing a comparison of the two sets by Tabar pattern and Boyd class (see Figs. 4 and 5).

2.1 Line Operator

A study of various methods for detecting linear structures in mammograms [10] showed that Dixon and Taylor's line operator [9] is more accurate than other methods. As such, the line operator was used in our experiments. The method produces a measure of line strength and orientation for each pixel in an image.

The line orientation is determined by calculating the mean pixel brightness of a line of pixels running through the target pixel at a range of orientations. The orientation with the largest mean brightness is taken to be the line orientation. The line strength, S, is then given by

$$S = (L - N), \tag{1}$$

where L is the mean brightness of the line of pixels, and N is the mean brightness of a similarly orientated square of pixels.

Our experiment used a line length of five pixels and twelve orientations as suggested by earlier work [10].

A multi-scale approach was used in order to detect lines of a range of thicknesses and the resultant images were combined to produce line strength values for pixels at the original scale. Scaling of the images was achieved firstly by blurring the image using a 3x3 Gaussian kernel and subsequently by subsampling to provide a resultant image of half the width and height of the original. Our approach comprised processing with the line operator at three scales, since this appeared to produce the most reasonable output for the images under examination.

Finally, the pixel line strengths were thresholded to remove background texture. Using a line length of 5, the measures of line strength fall in the theoretical range of $0 - 204$, however the results showed that most (if not all) pixels had line strength values in the range $0 - 30$. A range of threshold values were chosen experimentally, and two values (4/204 and 6/204) were used for our analysis as they removed most background noise whilst maintaining most of the linear structure information (see Fig. 1 (c, d)).

3 Results

Initial results focussed on the set of all images. The relative above threshold linearities for the various Tabar patterns and Boyd SCC classes are shown in Figs. 2 and 3 respectively. These graphs provide an overview of the difference between the patterns and classes, and for more detailed analysis, Mann-Whitney tests were performed on each pair of Tabar patterns and Boyd SCC classes. These are shown in Tables 1 and 2 respectively. Parametric tests, such as analysis of variance (ANOVA) tests were not used because the test data did not fulfill the necessary assumptions, however the relatively small number of classes made pairwise Mann-Whitney tests possible.

The Mann-Whitney test results provide an indication as to whether there was a statistically significant difference between two classes of images in our data set. A significant difference would mean that it is possible to reliably distinguish between different classes of images.

As mentioned, two threshold values were used in our analysis (4/204 and 6/204). This is to demonstrate the effects of varying the threshold level and thus including more or less of the linear structure information.

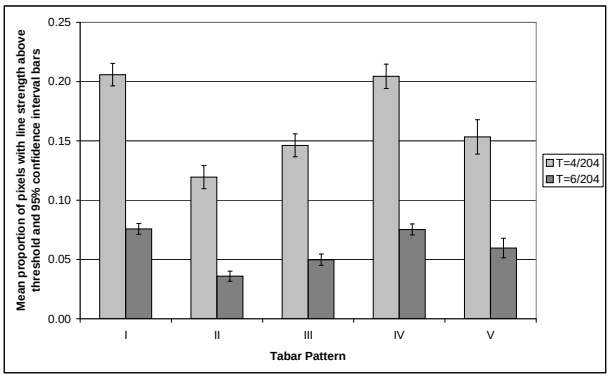


Figure 2. Graph showing the mean proportion of pixels with line strengths above each threshold T and 95% confidence intervals for images of each **Tabar** pattern.

Tabar	II	III	IV	V
I	0.0000	0.0000	0.7938	0.0000
II		0.0003	0.0000	0.0002
III			0.0000	0.3597
IV				0.0000

$T = 4/204$

Tabar	II	III	IV	V
I	0.0000	0.0000	0.9069	0.0059
II		0.0001	0.0000	0.0000
III			0.0000	0.0336
IV				0.0050

$T = 6/204$

Table 1. The p-values obtained by Mann-Whitney tests on each combination of **Tabar** patterns for each threshold T. Results not significant at $\alpha = 0.05$ are shaded.

The results of the analysis by Tabar pattern (see Fig. 2, Table 1) demonstrate the ability to reliably distinguish between patterns II–III (low risk) and patterns IV–V (high risk) at a threshold of 6/204, with low risk pattern I being indistinguishable from high risk pattern IV. The results at a threshold of 4/204 are less promising, since the low risk pattern III becomes indistinguishable from high risk pattern V.

The results of the analysis by Boyd SCC class (see Fig. 3, Table 2) differ somewhat, and it is clear that the proposed method is able to distinguish between classes 1–3 and classes 4–6 and both thresholds, with a general trend through classes 1–4/5 indicating that a greater linear density is indicative of a greater risk. The three lower risk classes are each distinguishable from all other classes, whilst the three higher risk classes are distinguishable from the three lower risk classes, but are indistinguishable from one another at $T = 6/204$. At $T = 4/204$ classes 4 and 6 become distinguishable from one another.

Following initial analysis, we investigated whether removing the abnormal images from the data set had an effect on results. The results of this analysis by Tabar pattern and Boyd SCC class are show in Figs. 4 and 5, respectively. The results show that there is no significant difference between the above threshold linearity in normal and abnormal images of each pattern or class.

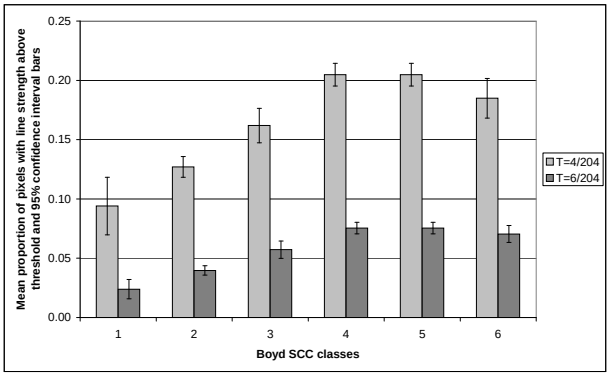


Figure 3. Graph showing the mean proportion of pixels with line strengths above each threshold T and 95% confidence intervals for images of each **Boyd SCC** class.

Boyd	**2**	**3**	**4**	**5**	**6**
1	0.0332	0.0023	0.0001	0.0001	0.0008
2		0.0002	0.0000	0.0000	0.0000
3			0.0000	0.0000	0.0362
4				0.6693	0.0287
5					0.0833

$T = 4/204$

Boyd	**2**	**3**	**4**	**5**	**6**
1	0.0155	0.0009	0.0001	0.0000	0.0004
2		0.0001	0.0000	0.0000	0.0000
3			0.0000	0.0000	0.0031
4				0.7621	0.4828
5					0.6612

$T = 6/204$

Table 2. The p-values obtained by Mann-Whitney tests on each combination of **Boyd SCC** classes at each threshold T. Results not significant at $\alpha = 0.05$ are shaded.

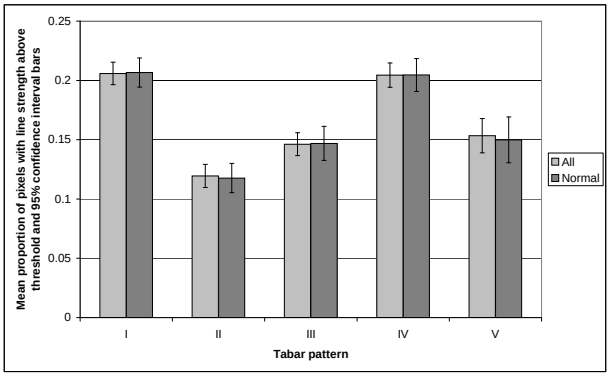


Figure 4. Graph showing the mean above threshold linearity and 95% confidence intervals for all images and the subset of normal images at each **Tabar** pattern.

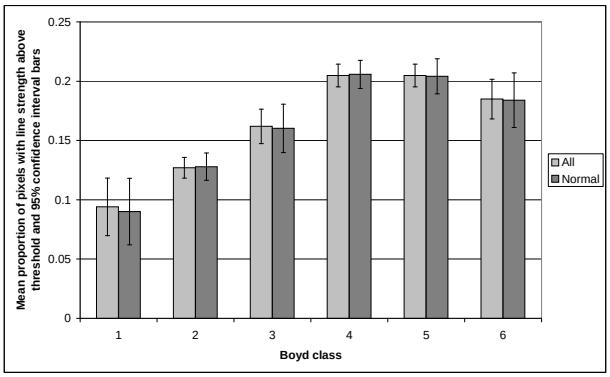


Figure 5. Graph showing the mean above threshold linearity and 95% confidence intervals for all images and the subset of normal images at each **Boyd** class.

4 Discussion and Conclusions

Whilst the proposed approach is simplistic, the results are promising and the analysis by Boyd class demonstrates a clear ability to automatically distinguish between lower risk (classes 1–3) and higher risk mammograms (classes 4–6).

The analysis by Tabar pattern is interesting. The results for pattern V were unexpected. Tabar suggests that the linear density in pattern V mammograms should be very low [5], whereas our results indicate a relatively high linear density for this pattern. However, a possible explanation might be that the line operator enhances linear structures even in dense tissue (see Fig. 1) and as such might result in a high proportion of linear structures, whereas under the Tabar classification this area might be assigned to one of the alternative mammographic building blocks, such as fibrous tissue. It is intended in future work to segment and mask the fibrous tissue prior to application of the line detector.

In addition, the results for Tabar pattern I mammograms, which demonstrate a high linear density, do not correlate with the results of low risk Boyd or BIRADS classes, which show low linear densities. Other studies have found that Tabar's patterns do not correlate well with other risk assessment models [11]. The principal anomaly is the low risk pattern I. The mammograms in our test set belonging to this pattern do not easily correlate with a particular Boyd class, instead being spread amongst Boyd classes 2-5, with the majority seeming to belong to the high risk classes 4-5 [11].

We see in the analysis by Boyd class that the highest class (6) shows a slightly lower linear density than the classes immediately below it. At a threshold of 6/204 this decline is only slight and there is no significant difference between the linear densities of classes 4–6 (see Table 2). The density of class 6 is also significantly higher than all three of the low risk classes. The decline is accentuated when a threshold of 4/204 is taken, where we see a significant difference between classes 4 and 6. This indicates a dependence on the threshold value and part of our future work will include an investigation in to a more principled approach to determine the threshold.

The results of a comparison between the whole data set and the subset of normal images shows no significant differences between the two sets. This is to be expected as the development of an abnormality is unlikely to increase or decrease the quantity of linear structures (with the exception of spiculations). It is possible, however, that it may affect the distribution of the linear structures.

In summary, the proposed approach is promising but simplistic in that it considers only the density of linear structures and does not take in to account information relating to their distribution. Since several risk assessment models are based on the parenchymal patterns in the breast [1, 3, 5], it is intended for further work to investigate whether the distribution of linear structures in addition to their density is related to mammographic risk, and whether this information can be used to improve risk assessment classification.

References

1. J. N. Wolfe. "Risk for breast cancer development determined by mammographic parenchymal pattern." *Cancer* **37(5)**, pp. 2486–2492, 1976.
2. T. M. Kolb, J. Lichy & J. H. Newhouse. "Comparison of the performance of screening mammography, physical examination, and breast us and evaluation of factors that influence them: An analysis of 27,825 patient evaluations." *Radiology* **225(1)**, pp. 165–175, 2002.
3. L. Tabar & P. B. Dean. "Mammographic parenchymal patterns. risk indicator for breast cancer?" *Journal of the American Medical Association* **247(2)**, pp. 185–189, 1982.
4. I. T. Gram, E. Funkhouser & L. Tabar. "The Tabar classification of mammographic parenchymal patterns." *European Journal of Radiology* **24(2)**, pp. 131–136, 1997.
5. L. Tabar, T. Tot & P. B. Dean. *Breast Cancer - The Art and Science of Early Detection with Mammography.* Georg Thieme Verlag, Stuttgart, 2005.
6. N. F. Boyd, J. W. Byng, R. A. Jong et al. "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study." *Journal of the National Cancer Institute* **87**, pp. 670–675, 1995.
7. American College of Radiology. *Illustrated Breast Imaging Reporting and Data System*. American College of Radiology, third edition, 1998.
8. E. M. Hadley, E. R. E. Denton & R. Zwiggelaar. "Mammographic risk assessment based on anatomical linear structures." *Lecture Notes in Computer Science* **4046**, pp. 626–633, 2006.
9. R. N. Dixon & C. J. Taylor. "Automated asbestos fibre counting." *Institute of Physics Conference Series* **44**, pp. 178–185, 1979.
10. R. Zwiggelaar, S. M. Astley, C. R. M. Boggis et al. "Linear structures in mammographic images: Detection and classification." *IEEE Transactions on Medical Imaging* **23(9)**, pp. 1077–1086, 2004.
11. I. Muhimmah, A. Oliver, E. R. E. Denton et al. "Comparison between Wolfe, Boyd, BI-RADS and Tabar based mammographic risk assessment." *Lecture Notes in Computer Science* **4046**, 2006.

Simulation of local and global atrophy in Alzheimer's disease studies

Oscar Camara[a], Martin Schweiger[a], Rachael I Scahill[b],
William R Crum[a], Julia A Schnabel[a], Derek LG Hill[a] and Nick C Fox[b] *

[a]Centre for Medical Image Computing (CMIC), Department of Medical Physics and Bioengineering,
University College London, Malet Place Engineering Building, London WCE1 6BT, UK
[b]Dementia Research Centre, Institute of Neurology,
University College London, Queen Square, London WC1N 3BG, UK

Abstract. We propose a method for atrophy simulation in structural MR images based on finite-element methods, providing data for objective evaluation of atrophy measurement techniques. The modelling of diffuse global and regional atrophy is based on volumetric measurements from patients with known disease and guided by clinical knowledge of the relative pathological involvement of regions. The consequent biomechanical readjustment of structures is modelled using conventional physics-based techniques based on tissue properties and simulating plausible deformations with finite-element methods. Tissue characterization is performed by means of the meshing of a labelled brain atlas, creating a reference volumetric mesh. An example of simulated data is shown and a visual classification protocol has been used by experts to rate real and simulated scans according to their degree of atrophy. First results demonstrate the potential of the proposed methodology.

1 Introduction

The pattern of cerebral atrophy on magnetic resonance imaging (MRI) is currently used to help differentiate the degenerative dementias while the rate of atrophy may be used to detect early disease and potentially to assess disease-modifying effects of therapies. Algorithms referred to as computational anatomy [1] are computerized approaches that offer automated or semi-automated solutions for MRI analysis, including quantification of cerebral atrophy. There is a lack of a gold standard against which to judge these methods or to help refine them. There have been limited attempts [2, 3] to model the complex effects on regional structural volumes and structural configuration that are a consequence of neurodegeneration. The main drawback of these approaches is that they do not take into account the interrelation of different tissue types.

We present a method for the simulation of atrophy in MR images of the brain in which atrophy is simulated in different tissue compartments and/or in different neuroanatomical structures with a phenomenological model. This model applies a differential pathological burden in different brain regions. Once volume changes have been defined in particular structures and tissues, the overall impact on brain shape is determined by a conventional physics-based biomechanical model. The consequent readjustment of structures is modelled using techniques based on tissue properties and simulating plausible deformations with finite-element methods (FEM). This validation framework was proposed by Schnabel et al. [4], for the analysis of MR mammography data.

The proposed technique can be divided into three main steps: meshing of a reference labelled brain atlas; its introduction into a FEM solver that will generate the simulated deformations; and the application of such deformations to the grey-level version of the brain atlas.

2 Mesh generation

In this work, we have combined information from two well-known brain atlases to generate the reference labelled image and mesh used by the FEM solver: the Montreal Neurological Institute (MNI) Brainweb [1]; and the International Consortium of Brain Mapping (ICBM) [2] atlases.

For our purposes, we have taken into account the tissue labels from the MNI atlas, including the cerebrospinal fluid (CSF), the grey matter (GM) and the white matter (WM). We have also incorporated labels for the hippocampi and the lateral ventricle, obtained from a semi-automatic segmentation by clinical experts [5]. In addition, we have incorporated the cerebral lobe labels from the ICBM Probabilistic atlas using an affine registration technique in order to place this atlas into the same geometrical space as the MNI one. Finally, two additional labels from the ICBM Template atlas

* Corresponding author: o.camara-rey@ucl.ac.uk
[1] http://www.bic.mni.mcgill.ca/brainweb/
[2] http://www.loni.ucla.edu/ICBM/

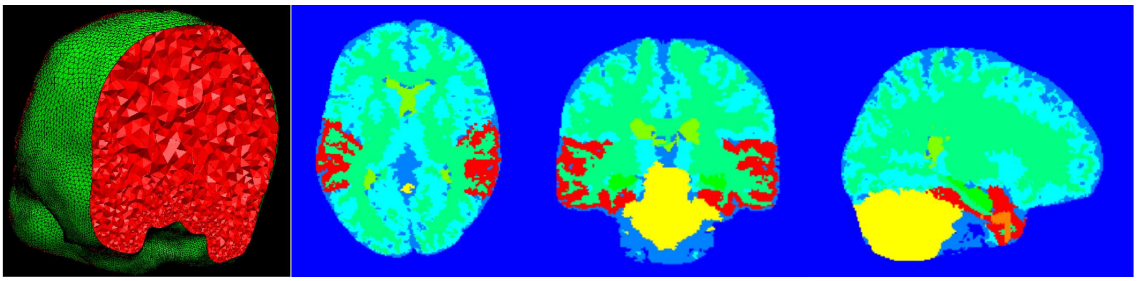


Figure 1. Meshing stage. Cutting plane of the volumetric finite-element mesh (left) and slices of the rasterization image generated from the labelled mesh, in which some of the used labels are shown.

have been included: a label covering the subtentorial structures to specify boundary conditions of the FEM; and a label for the entorhinal area (ERA) because of its implication in Alzheimer's disease. Some of these labels are illustrated in Figure 1.

The first phase of the meshing consists of the generation of a surface mesh from the intracranial surface using the classical marching cubes [6] algorithm, followed by some post-processing steps (decimation, smoothing and selection of biggest connected component). A 3D finite-element (4-noded tetrahedra) mesh was then obtained from the intracranial mesh surface using the NETGEN [3] openSource software. We generated a tetrahedral mesh with very fine resolution (163765 nodes and 868404 elements, with 61132 surface elements). The final step is the labelling of the 3D mesh using the labelled voxel atlas, in which a set of particular biomechanical properties is assigned to each tetrahedron through a rasterization procedure [7].

3 Finite-element deformation model

3.1 Model

The deformation model employs a linear elastic finite-element method based on the freely available TOAST package [4]. In this work we considered structural displacements induced by atrophy that are small in relation to the total brain volume, so that the use of a linear model is justified.

The implementation of the FEM model follows a standard approach, whereby each tetrahedral mesh element is assigned a set of elastic material properties represented by an elasticity matrix D. In the case of isotropic [5] elastic deformations, D is symmetric and can be expressed in terms of two parameters, usually given as Young's modulus E and Poisson's ratio ν. The elastic coefficients are assumed time-invariant. The deformation of the mesh is induced by assigning an isotropic thermal expansion coefficient $\alpha^{(i)}$ to each element i, and simulating a global temperature change ΔT. The resulting isotropic thermal expansion enters the description of elastic deformation in the form of an initial element strain,

$$\varepsilon_0^{(i)} = \{\alpha^{(i)}\,\Delta T, \alpha^{(i)}\,\Delta T, \alpha^{(i)}\,\Delta T, 0, 0, 0\}, \tag{1}$$

where the relation between stresses σ and strains ε is given by

$$\sigma^{(i)} = \mathsf{D}^{(i)}(\varepsilon^{(i)} - \varepsilon_0^{(i)}) + \sigma_0^{(i)} \tag{2}$$

Assembling all element contributions of the mesh leads to the linear system

$$\mathsf{K}\mathbf{u} + \mathbf{f} + \mathbf{f}' = \mathbf{r}, \qquad K_{ij} = \int_V \mathsf{B}_i^T \mathsf{D} \mathsf{B}_j dV \tag{3}$$

with stiffness matrix K. In an n-noded element, B is a $6 \times 3n$ strain displacement matrix $\mathsf{B} = \{\mathsf{B}_i\}$, and

$$\mathbf{f}' = \int_V \mathsf{B}^T \mathsf{D} \varepsilon_0 dV, \tag{4}$$

contains the volume forces arising from the initial thermal strain, $\mathbf{f}$ combines all other surface and volume force terms, $\mathbf{r}$ defines explicit displacements, and $\mathbf{u}$ is the vector of nodal displacements.

[3] http://www.hpfem.jku.at/netgen/

[4] http://www.medphys.ucl.ac.uk/∼martins/toast/index.html

[5] Future work will investigate the use of anisotropic elastic properties, especially relevant in highly fibrous tissues such as the white matter.

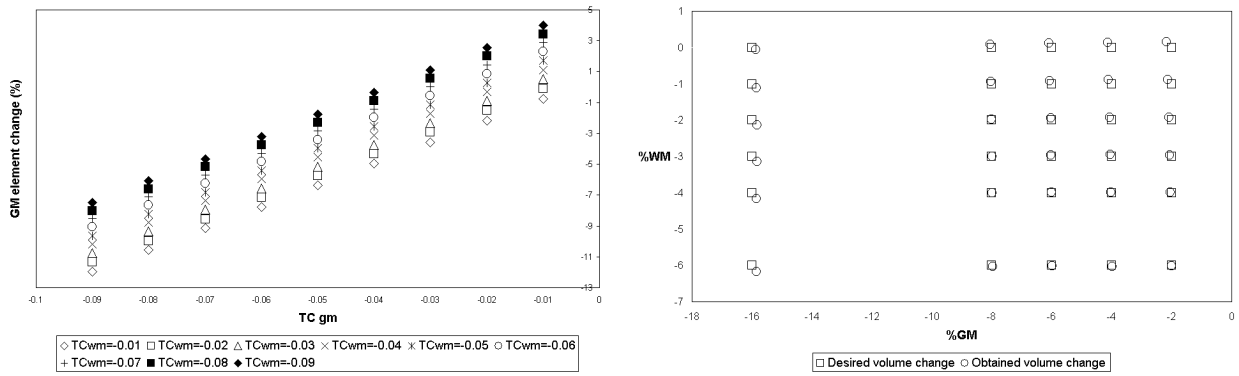


Figure 2. Relation between thermal coefficients and desired volume change. Left: Percentage of GM element change when varying the thermal coefficients of the GM and WM (TCgm and TCwm respectively). Right: Relation between desired and the obtained volume change of GM and WM.

3.2 Boundary conditions and mechanical properties of brain tissue

The surface of the mesh was assumed to coincide with the inner surface of the skull, and homogeneous Dirichlet boundary conditions were introduced using a Payne-Irons method to suppress the displacements of boundary nodes. The same strategy was applied to the mesh nodes corresponding to the subtentorial structures since atrophy induced by dementia is normally restricted to the supratentorial area.

The choice of the elastic properties is not critical in the model presented here, because the boundary conditions enforce a constant global volume, while the relative changes in tissue region volumes are driven by the ratio of thermal coefficients. Furthermore, the mechanical properties of brain tissue are directly relevant to how the brain responds to an external force but are much less relevant to the response of the brain to diffuse cell damage and death that results in active structural readjustment. Therefore, we have adopted the shear modulus values proposed by McCracken et al. [8]: 12KPa for WM and 8KPa for GM (correspondingly, Young's modulus of 34.8KPa for WM and 23.2KPa for GM). We consider the remaining structures, except the CSF, as being composed of an equal percentage of GM and WM, then their material properties are the average of the ones corresponding to the GM and WM ($E = 29.0$KPa). For the Poisson's ratio, we have chosen a value of 0.45 for each brain tissue, except for the CSF, assumed to be a very soft and compressible tissue, with low Young's modulus ($E = 0.1$KPa) and Poisson's ratio ($\nu = 0.05$).

3.3 Thermal coefficient computation

Given a segmentation of domain Ω into N tissue types or anatomical structures $\Omega_j,\ \Omega = \cup_j \Omega_j,\ j = 1..N$, we assume the thermal coefficient α to be constant within each region Ω_j, so that α is expressed in a piecewise constant basis, $\alpha(\mathbf{r}) = \alpha_j,\ \mathbf{r} \in \Omega_j$. Once the material properties have been defined, we need to estimate the thermal coefficients α_j that will provide a mesh deformation consistent with the desired volume change ΔV_j in each tissue segment Ω_j. This volume change will provide the ground-truth information needed for validation. ΔV_j is computed by integrating the differences in element volumes between the original and displaced meshes over each region Ω_j.

The method of recovering the thermal coefficients for a given set of target region volume changes is based on a linear inversion technique. The relationship between changes of α_j and ΔV_i in regions i and j is expressed by the Jacobian matrix

$$\mathsf{J}_{ij} = \left[\frac{\partial V_i}{\partial \alpha_j}\right], \tag{5}$$

which we compute with a finite difference scheme, by explicitly perturbing each α_j. Now we form the linear system $\mathsf{J}[\alpha] = [\Delta V]$, which, taking into account the boundary conditions, is quadratic of dimension $(T-1) \times (T-1)$ and can be solved by a standard LU decomposition. As illustrated in Figure 2, the assumption of linearity is appropriate. In this figure, we also show the good agreement between the desired and the obtained volume change for a set of 30 simulations (mean difference of 0.35 ± 0.41 and 0.14 ± 0.08 for GM and WM, respectively), confirming the appropriateness of the linear model.

Table 1. Volume changes (% with respect to the baseline) simulated in the longitudinal example. MTL: Medial Temporal Lobe. Most significant volume changes in the AD patients in each timepoint are marked in bold.

Structures	Normal aging			Alzheimer's disease		
	t1	t2	t3	t1	t2	t3
Cortical CSF	1.90	3.76	5.66	1.98	5.27	**16.82**
Ventricles	2.18	4.31	6.49	2.23	**5.82**	**19.34**
Grey Matter	-0.65	-1.28	-1.92	-0.65	-1.30	**-5.31**
White Matter	-0.17	-0.32	-0.49	-0.16	-0.32	**-1.33**
Hippocampi	-1.40	-2.79	-4.21	**-3.33**	**-6.39**	**-9.31**
Subtentorial	0.02	0.03	0.05	0.02	0.04	0.13
Entorhinal Area	-2.91	-5.80	-8.67	**-5.11**	**-9.60**	**-13.68**
MTL Grey Matter	-0.63	-1.28	-1.92	-0.65	**-4.68**	**-8.59**
Whole Brain	-0.36	-0.71	-1.08	-0.38	-1.01	**-3.20**

4 Warping

In order to simulate atrophy in the atlas image, we need to use interpolation techniques to generate a displacement field at each voxel of the image. We do this by weighting the displacement vectors of the nodes in the element containing a given voxel by the element linear shape function [9], and then the new image intensities can be interpolated using a truncated sinc interpolation kernel. We have adopted a different interpolation strategy outside the mesh, where no deformation information is available from the FEM model. Here we employ a scattered data interpolation technique based on multilevel B-splines, as in Schnabel et al. [4].

5 Simulation example

The illustrative example shown in this section is focused on Alzheimer's disease, as it is the most common cause of dementia. Therefore, the longitudinal example is based on the generation of a set of simulated images approximately mimicking the distribution of regional change in AD according to the severity of the disease. Three different atrophy simulated images are generated: in the first one ($t1$), only changes in the hippocampi and the entorhinal area are applied; the second image ($t2$) also includes GM changes in the medial temporal lobe and an expansion of the lateral ventricles; finally, a global reduction of the whole brain is applied to generate the third simulated image ($t3$). In addition, we have generated a corresponding set of simulated images at the three timepoints representing volume change induced by normal aging. We chose a range of values (see Table 1) which included a range of cerebral volume losses, as a percentage of baseline, that is consistent with published rates of atrophy and study durations [10] of 6-24 months. Figure 3 shows the difference images between the normal aging and the AD simulated images at each timepoint.

6 Visual classification protocol

The aim of this modelling is to generate simulated atrophy for the validation of techniques for the quantification of local and global atrophy. Therefore, an experiment was designed in which a visual classification protocol was used by experts to rate real and simulated MRI data of controls and patients with atrophy. We could then study whether experts consistently rate real and simulated data in a different way. The experiment uses two matched cohorts of 20 healthy elderly controls and 20 individuals with a diagnosis of probable Alzheimer's disease, together with nine atrophy simulated images. T1-weighted volumetric MR images were acquired on a 1.5 Tesla Signa unit (General Electric, Milwaukee).

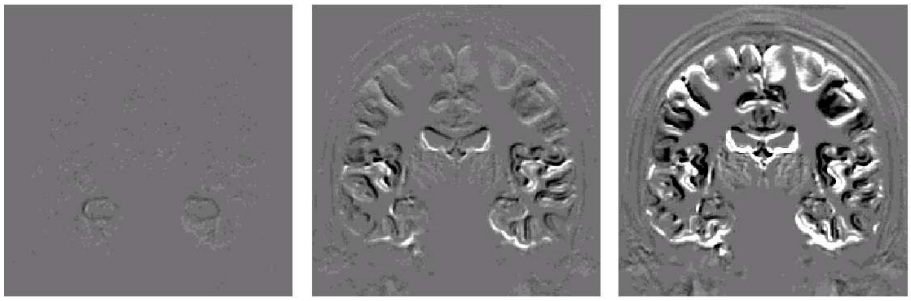


Figure 3. Results of the longitudinal simulation example. Coronal slices of the difference images between the normal aging and the AD simulated images at each timepoint (from left to right, $t1$, $t2$, $t3$).

A mesh warping [11] technique is employed to put the volumetric mesh presented in Section 2 into the anatomical space of nine of the controls, thus generating nine control-specific meshes. Thereafter, these meshes are introduced into the FEM solver, together with information about the desired volume change, which in this work we have obtained from volumetric difference estimates between both cohorts that have been previously computed using manual tracing techniques. The output of the FEM solver is a set of deformations that are applied to the control images to generate a group of atrophy simulated images equivalent to the cohort of mild AD subjects in terms of regional atrophy. The three sets of real and simulated data are then presented to the neurologists, who were blinded to scan information and were asked to classify each scan on a five point scale, with one being definite control, and five being definite AD.

The mean sensitivity and specificity for visual assessment of true AD and control scans was 90% and 77% respectively. The inter-rater agreement was moderate, with a kappa score of 0.52. The mean score for each scan category is (mean$\pm$STD): 1.95$\pm$1.42 for controls; 4.23$\pm$3.19 for real AD; 3.19$\pm$1.36 for simulated AD. The simulated scans were not consistently classified as AD, but our results show that even true AD cases may be misidentified in one in ten cases. This may be due to the large inter-subject variation which exists when these subject groups are compared cross-sectionally. Nevertheless, the visual classification protocol suggests that the methodology produced scans leading to the same trend in classification in a cross-sectional study than when using real data. From this observation, we can conclude that an expert may not always be able to discern between the simulated scans and real data, thus implying that our simulation produces results with reasonable degree of realism.

7 Conclusions

We have proposed a method to generate valuable gold standard data for the testing and objective quantitative validation of atrophy measurement techniques. Our technique uses expert clinical knowledge of the rate and spatial distribution of atrophy in dementia to create a phenomenological model of the appearance of atrophy in MRI based on finite-element methods. The model includes structural remodelling concomitant with atrophy. Future work will focus on the generation and dissemination of several cohorts of data simulating the pattern of changes induced in the human brain by different types of dementia. The simulated scans will be made available to the wider research community.

Acknowledgements

OC acknowledges support of the EPSRC GR/S48844/01. WRC and MS acknowledge support of the EPSRC GR/N14248/01 and the UK Medical Research Council Grant No. D2025/31. JAS acknowledges support of the EPSRC GR/S82503/01. RIS and NCF acknowledge support from the UK Medical Research Council, G90/86 and G116/143 respectively.

References

1. J. Ashburner, J. Csernansky, C. Davatzikos et al. "Computer-assisted imaging to assess brain structure in healthy and diseased brains." *Lancet Neurology* **2**, pp. 79–88, 2003.
2. Z. Lao, D. Shen, Z. Xue et al. "Morphological classification of brains via high-dimensional shape transformations and machine learning methods." *Neuroimage* **21(1)**, pp. 46–57, 2004.
3. K. Chen, E. Reiman, G. Alexander et al. "An automated algorithm for the computation of brain volume change from sequential MRIs using an iterative principal component analysis and its evaluation for the assessment of whole-brain atrophy rates in patients with probable Alzheimer's disease." *Neuroimage* **22**, pp. 134–143, 2004.
4. J. Schnabel, C. Tanner, A. Castellano-Smith et al. "Validation of Non-Rigid Registration using Finite Element Methods." In *Information Processing in Medical Imaging (IPMI'01)*, volume 2082, pp. 183–189. 2001.
5. P. Freeborough, N. Fox & R. Kitney. "Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans." *Computer Methods and Programs in Biomedicine* **53**, pp. 15–25, 1997.
6. W. Lorensen & H. Cline. "Marching cube, a high resolution 3D surface reconstruction algorithm." In *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'87)*, volume 21, pp. 163–169. 1987.
7. M. Sermesant, C. Forest, X. Pennec et al. "Deformable biomechanical models: Application to 4D cardiac image analysis." *Medical Image Analysis* **7**, pp. 475–488, 2003.
8. P. McCracken, A. Manduca, J. Felmlee et al. "Mechanical Transient-Based Magnetic Resonance Elastography." *Magnetic Resonance in Medicine* **53(20)**, pp. 628–639, 2005.
9. A. J. Davies. *The Finite Element Method: A First Approach*. Oxford University Press, 1980.
10. C. Jack, Jr., M. Shiung, J. Gunter et al. "Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD." *Neurology* **62**, pp. 591–600, 2004.
11. O. Camara, W. Crum, J. Schnabel et al. "Assessing the quality of Mesh-Warping in normal and abnormal neuroanatomy." In *Medical Image Understanding and Analysis (MIUA'05)*. 2005.

Simulation of Acquisition Artefacts in MR scans: Effects on Automatic Measures of Brain Atrophy

Beatrix I Sneller[a], Oscar Camara[a], Gerard R Ridgway [a], Ellen Garde[b], Nick C Fox[b] and Derek LG Hill[a]

[a]Centre for Medical Image Computing (CMIC), UCL, [b]Dementia Research Centre, UCL, Institute of Neurology

Abstract. Automatic algorithms in conjunction with longitudinal MR brain images can be used to measure cerebral atrophy, which is particularly pronounced in several types of dementia. An atrophy simulation technique has been devised to facilitate validation of these algorithms. To make this model of atrophy more realistic we simulate acquisition artefacts which are common problems in dementia imaging: motion (both step and periodic motion of increasing degrees of rotation) and pulsatile flow artefact. Artefacts were simulated by combining different portions of k-space from various modified image. The original images were 7 MR scans of healthy elderly controls, each of which had two levels of simulated atrophy. We investigate the effect of the simulated acquisition artefacts in atrophy measurements provided by an automatic technique, SIENA.

1 Introduction

Structural magnetic resonance imaging (MRI) can be used to estimate the progression of neurodegenerative disease since brain atrophy is a measurable feature of the most common forms of dementia such as Alzheimer's disease (AD). Several computerized techniques [1-4] for the quantification of cerebral atrophy in MR scans are available in the literature. Nevertheless, there is a lack of consensus on the most appropriate method to be used since standardized test datasets with known atrophy rates are not available.

We have described [5] a method based on the combination of expert knowledge of clinical changes in brain with finite-element methods, providing valuable groundtruth data for the objective validation of atrophy measurement techniques. This technique, however, did not take into account the common presence of artefacts in the MR scans due to imperfections in the acquisition, which may have a strong impact on the performance of the subsequent atrophy quantification techniques. For instance, movement during the acquisition of the edge of k space (domain in which raw MRI data is stored. The Fourier transform of this data is the MR image) lead to a ringing artefact from the scalp [6], global movement during the central portion of k-space would generate obvious ghosts and blur image intensities, or flow artefacts may appear in the medial temporal lobes due to the pulsatile blood flow in the carotid arteries. All these artefacts have the potential to severely degrade quantitative MR-based analysis such as measurement of atrophy.

The main goal of this paper is to present the techniques we are developing to simulate acquisition artefacts in order to make the generated ground-truth data more realistic. Furthermore, these simulations will allow the assessment of post-processing artefact correction techniques [6, 7, 8] and their incorporation in atrophy measurement methods.

One of the most common causes of acquisition artefacts is subject motion. Rotational motion artefacts were ranked by Blumenthal et al. [9] to study their effects on a grey matter quantification technique and simulated by Preboske et al. [10] to investigate their impact on a semi-automatic atrophy measurement method, the boundary shift integral [1]. Nevertheless, in some particular brain regions, other factors can be as important as subject motion. For instance, the main cause of image degradation in the temporal lobe, a region with a crucial role in Alzheimer's disease, is pulsatile flow artefact from carotid arteries. In clinical studies related to dementia, out of 837 subjects, we have visually identified 105 (12.5%) of the subjects' images with artefacts located in the temporal lobes, which can be classified according to their cause: 68 (64.8%) with flow artefact; 16 (15.2%) with motion-related artefact; 4 (3.8%) with susceptibility artefacts; and 17 (16.2%) images with artefacts induced by other factors.

We have developed techniques to simulate the two most significant causes of acquisition artefacts, a rotational motion during the scan, and a pulsatile motion. We have then applied SIENA [2], an automatic technique for the estimation of atrophy rates, to different sets of atrophy simulated data with the presence of acquisition artefacts.

2 Method

Beatrix Sneller, University College London, Dept. of Computer Science, Gower Street, London, WC1E 6BT
b.sneller@ucl.ac.uk, o.camara-rey@ucl.ac.uk, Gerard.Ridgway@ucl.ac.uk, nfox@dementia.ion.ucl.ac.uk, derk.hill@ucl.ac.uk

The generation of ground-truth data from an MR scan for the evaluation of atrophy measurement techniques can be divided in two stages: simulation of atrophy and simulation of acquisition artefacts.

2.1 Simulation of Atrophy

The method for atrophy simulation can be divided into four phases: meshing of a labelled brain atlas; adaptation of the generated mesh to an individual MR scan; introduction of the individual-specific mesh into a FEM solver that will generate the simulated deformations; and the application of such deformations to the MR scan. More details about this method can be found in [5]. Figure 1 shows an example of atrophy simulation.

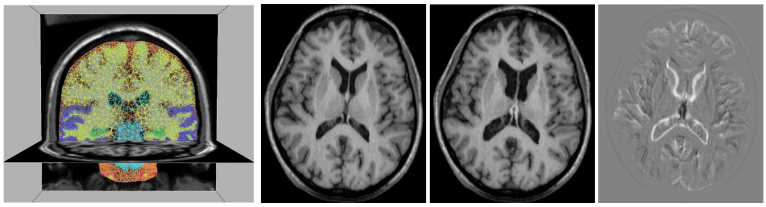


Figure 1. From left to right: mesh, baseline, atrophy simulated scan and difference image

2.2 Simulation of Acquisition Artefacts

2.21 Rotational motion

Rotational motion during a scan causes inconsistencies in the sampled k-space signal that result in blurring and ghosting in the image domain. The method we developed to simulate rotational motion is the following. A motion-free 3D image volume (top left of Fig. 2) is converted into a complex image by treating the modulus image as the real part, and setting the imaginary component to zero. This complex image is then rotated by a set amount, in this case normally 5 degrees, about an axis through the internal auditory meatus, simulating a slight nodding of the patient during the scan (bottom left of Fig. 2). Bicubic interpolation was used in the rotation, and voxels entering the volume were assigned the value zero. The original and rotated images were then both Fourier transformed into k-space (second column of Fig. 2). A motion-corrupted k-space was subsequently generated by assembling portions of the k-spaces corresponding to the un-rotated and rotated images (third column of Fig. 2). The motion-corrupted k-space was then transformed back to the image domain (final image of Fig. 2) to give the motion corrupted volume. The k-space samples were combined as sagittal planes representing a sagittal MR acquisition orientation. Transverse or coronal planes could also have been simulated in these 3D volumes by appropriately assembling the final k-space.

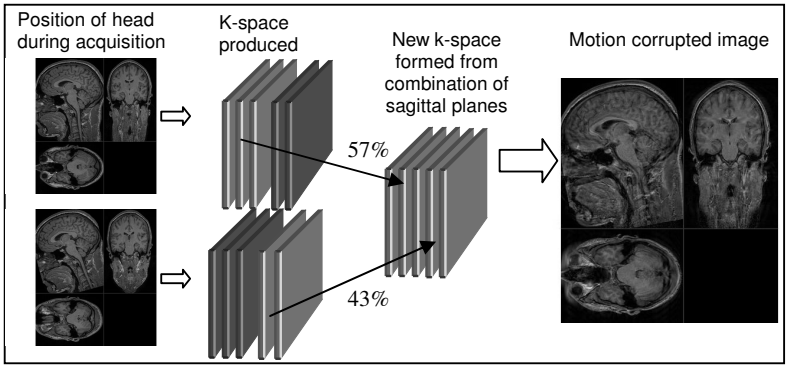


Figure 2. Simulation of step rotational motion: motion is applied in the spatial domain to a motion-free image volume, and the k-spaces re-combined to generate a motion-corrupted image volume.

The appearance of the motion artefacts is influenced by which region (its distance to the centre of the k-space) of k-space is being sampled as the motion takes place, and whether there is a single step motion, or a periodic motion. Our technique allows complete freedom in the selection of the time of the motion in the scan, and for the examples in this paper, we simulated step motion by combining the first 57% of the un-rotated k-space, with 43% of the rotated k-space. We simulated the motion just after half way through the acquisition as the artefact is more noticeable the nearer to the centre of k-space it occurs. Periodic motion, representing a repeated nodding of the head, was simulated by generating the motion-corrupted k-space from portions of the un-rotated and rotated k-space, alternating between the two at 44%, 48%, 52%and 56%. Figure 3 shows an example of the images resulting from simulated step and

periodic rotational motion. One of the most troublesome causes of motion is step motion near the middle of the scan (eg: if someone coughs or moves suddenly). We have chosen to simulate this sort of motion in this work, but other types of motion such as slow drifts can also be simulated.

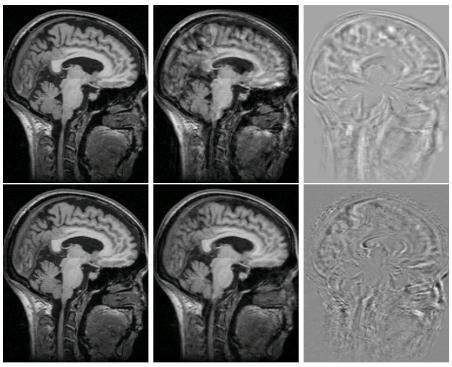


Figure 3. Top: periodic motion, bottom: step motion. From left to right: Atrophy simulated image, atrophy plus motion artefact simulated images, difference image.

2.22 Pulsatile Motion

Another important artefact in MR images is pulsatile motion, which typically leads to streaks across the image in the slow-phase encoding direction. These artefacts arise from major blood vessels in the field of view, such as the carotid artery where it passes through the carotid siphon, and result from a signal in the vessel that varies from shot to shot during the acquisition. Simulated pulsatile motion artefacts were added to the simulated data as follows (see also Figure 4). The 3D modulus images were converted to a complex image as described above, and four different coil views were obtained by multiplying the image by four separate coil sensitivity profiles. Because the pulsatility artefact is predominantly in the slowphase encode direction, each plane of k-space was treated as being acquired instantaneously (contrast being largely determined by the acquisition of the centre portion of k-space) and 1.73s apart from the next. A cylinder corresponding to the carotid artery was manually delineated from the image, and the intensity of this modulated according to a literature-based model of the flow in the carotid artery [11]. For each shot in the slow-phase encode direction, a different image was simulated by modulating the intensity of the carotid signal in the delineated cylinder. These images were transformed into the k-space, and one k-space plane from each image was combined to form a pulsatile-motion corrupted k-space based on an estimated heart rate of 60bpm. Because this is a one dimensional motion, it was implemented just for the lines containing the carotid, rather than the entire image, in order to reduce computational cost.

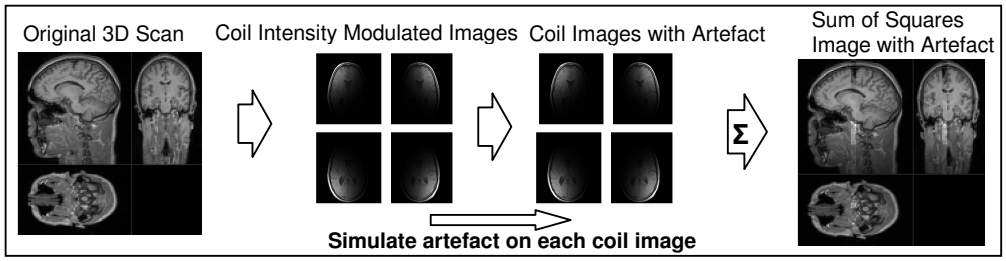


Figure 4. Overview of simulation of flow artefact

The resulting image, as can be seen in Figure 5 has a simulated flow artefact running right to left. This is because we simulated the phase encode direction in the left-right direction as would be normal for a sagittally acquired head volume. This is implemented in the simulation by using sagittal planes of k-space. However we could simulate any direction of acquisition which would alter the direction of the simulated artefact, and could also simulate more complex vessel geometries or flow properties.

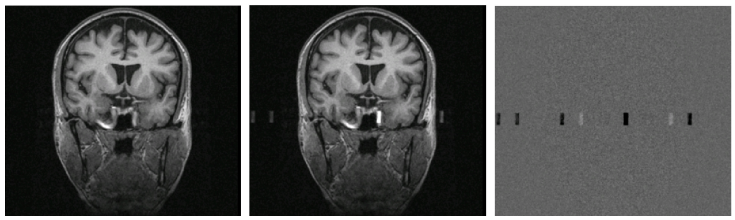


Figure 5. From left to right: Atrophy simulated image; atrophy and flow artefact

3 Experiment

3.1 Design

We have designed an experiment in order to investigate the effect of the simulated acquisition artefacts in atrophy measurements provided by an automatic technique, SIENA (Structural Image Evaluation, using Normalization, of Atrophy), proposed by Smith et al. [2]. Firstly, this method segments brain from non-brain tissue in the head, and estimates the outer skull surface. Then, it uses these results to register the two brains with the skulls constrained. These images are then segmented, and the atrophy is estimated from the movement of the segmentation boundaries. The output is the estimation of the percentage of brain volume change between the two scans, as well as a flow image in which regions of local change are shown, as can be seen in Figure 5. The authors of this technique claim a median absolute error of 0.2% of the whole brain volume change in a validation study.

In this work, we have employed 7 MR scans of healthy elderly controls, enrolled in a longitudinal research project at the Dementia Research Centre, UCL, Institute of Neurology, UK and who underwent serial MRI, clinical and neuropsychological assessment over a period of at least one year. T1-weighted volumetric MR images were acquired on a 1.5 Tesla Signa Unit (General Electric, Milwaukee) using a 256*256 matrix to provide 124 contiguous 1.5mm coronal slices through the head (acquisition parameters: time to repeat, 15ms; time to echo, 5.4ms; flip angle 15°, field of view 24*24cm). We have applied the atrophy simulation technique [6] to these 7 MR scans, generating two new images, *Atsim1* and *Atsim2*, for each subject scan (baseline) with different amounts of regional and global atrophy in several brain tissue compartments and structures, resulting in a dataset of 21 images. In this work, we only make use of the atrophy simulation's ground truth values for the whole brain volume change since SIENA only gives an estimation of the global atrophy rate. The mean and standard deviation of the simulated whole brain volume change percentages are the following: a reduction of 5.11% ± 0.23 for *Atsim1*; and a reduction of 10.84% ± 0.47 for *Atsim2*. The next phase of the experiment applies the methods described in Section 2 to the dataset of atrophy simulated images in order to simulate the acquisition artefacts detailed above. Specifically, we simulate step and periodic rotational motion (see Section 2.1) for the 21 available images and pulsatile motion (see Section 2.2) for a subset of 7 images (three different cases and their corresponding atrophy simulated images), due to the time-consuming manual segmentation of the carotid artery needed to simulate this artefact. Furthermore, we have obtained, for a single case and its associated *Atsim1*and *Atsim2* images, 10 different step and periodic rotational motion simulated images varying the severity of the artefact (i.e. instead of just using the original 5 degrees rotation, using rotations from 0° to 10° degrees in intervals of 1 degree). Finally, we run SIENA on the whole dataset of simulated images, aiming at studying the effect of acquisition artefacts on atrophy measurements.

3.2 Results

Results are summarized in Table 1. We have computed the difference (diff1) between the whole brain atrophy measurement calculated by SIENA and the gold standard, for the images that contain both simulated atrophy and simulated artefacts. We have also computed the difference (diff2) in SIENA results obtained with just simulated atrophy, and with simulated atrophy plus simulated artefact. The main difference between both measures is that "diff1" accumulates SIENA errors in the estimation of atrophy as well as errors induced by the simulation of acquisition artefacts, while "diff2" is independent of the atrophy simulation stage.

	No artefact	Periodic	Step	Flow
diff1	0.286±0.383	1.197±0.777	1.123±0.904	0.412±0.220
diff2	n/a	0.982±0.659	0.941±0.738	0.479±0.099

Table 1. Effect of simulated acquisition artefacts. diff1: difference between SIENA results and the gold standard for atrophy and artefact simulated images. diff2: difference in SIENA results between just simulated atrophy and with simulated atrophy plus simulated artefact. diff1 and diff2 are percentages of whole brain volume change.

Both types of difference measures suggest that the presence of artefacts in the images has a substantial impact on atrophy measurements obtained with SIENA. We can observe in Table 1 that rotational motion artefacts, either step or periodic, degrade the performance of SIENA more than flow artefacts. This is due to the limited region where pulsatile artefacts are present, mostly the temporal lobes, while SIENA only provides a global estimation of the whole brain volume change. It must be pointed out that the accuracy obtained with SIENA in atrophy measurements without artefacts (first column of Table 1) are consistent with results provided by the authors of this technique (accuracy of 0.2% of whole brain volume change). Figure 6 shows the impact on the error in the SIENA atrophy measurement as a function of magnitude of the rotational motion (rotation from 0° to 10° degrees in intervals of 1 degree). There is a strong correlation between the increase in atrophy measurement errors and the severity of the simulated artefact. The expected difference in whole brain volume change between AD and normal aging over a 12 month period is around

1.5%. Results in Figure 6 suggest that rotational motions larger than 5 degrees of rotation could induce artefacts around the same order of these expected differences.

4 Conclusions

We have presented the techniques we have developed to simulate the two most common acquisition artefacts in MR scans, rotational (both step and periodic) and pulsatile motion, in order to make ground truth data for atrophy measurement techniques more realistic. The methodology can be used to simulate many different magnitudes of artefacts, and can be applied to many different atrophy quantification methods.

Only one atrophy quantification method has been investigated. The impact is likely to be different with different algorithms, and especially for local atrophy measures (e.g.: from a non-rigid registration algorithm) or cortical thickness measurement methods. The value of the approach is in predicting the sensitivity of methods to motion artefact (potentially valuable in a power analysis used in the design of a clinical trial), and also in comparing the relative merits of alternative atrophy measuring algorithms in the presence of common image artefacts.

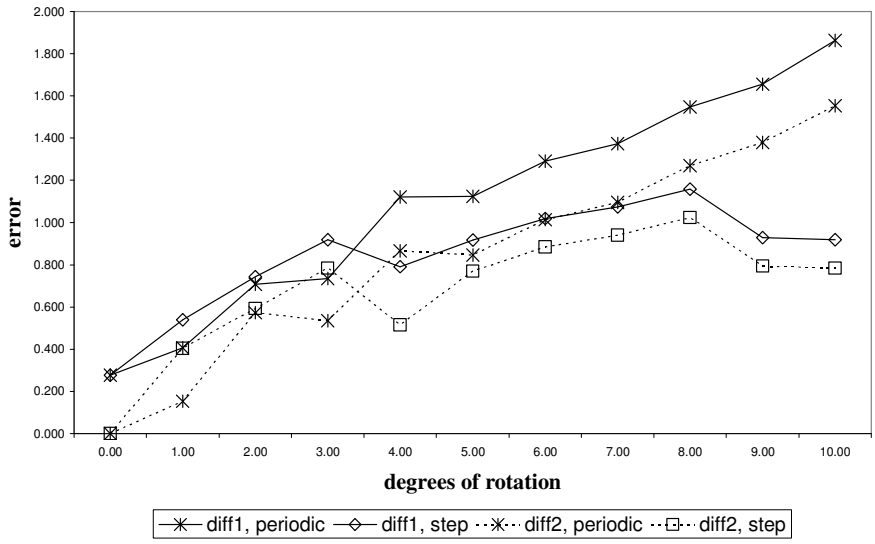


Figure 6. Effect on SIENA results, in percentage brain volume change, of varying the severity of the rotational simulated artefacts

References

1. Freeborough, P., Fox, N.: The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Transactions on Medical Imaging 16 (1997) 623–629
2. Smith, S., Stefano, N.D., Jenkinson, M., Matthews, P.: Normalized accurate measurement of longitudinal brain change. Journal of Computer Assisted Tomography 25 (2001) 466–475
3. Davatzikos, C., Genc, A., Xu, D., Resnick, S.: Voxel-based Morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. Neuroimage 14 (2001) 1361–1369
4. Freeborough, P., Fox, N., Kitney, R.: Modeling brain deformations in Alzheimer's disease by fluid registration of serial 3D MR Images. Journal of Computer Assisted Tomography 22 (1998) 838–843
5. Camara, O., Schweiger, M., Scahill, R.I.,Crum, W.R., Sneller, B.I., Schnabel, J.A, Ridgway, G.R., Cash, D.M., Hill, D.L.G., Fox N.C.: Phenomenological model of diffuse global and regional atrophy using Finite-Element methods. IEEE Transactions on Medical Imaging (2006). In press.
6. Howarth, C., Hutton, C., Deichmann, R.: Improvement of the image quality of T1-weighted anatomical brain scans. Neuroimage 29 (2005) 930–937
7. Atkinson, D., Hill, D.L., Stoyle, P.N., Summers, P.E., Clare, S., Bowtell, R., Keevil, S.F.: Automatic compensation of motion artifacts in MRI. Magnetic Resonance Medicine 41 (1999) 163–170
8. Manduca, A., McGee, K.P., Welch, E.B., Felmlee, J.P., Grimm, R.C., Ehman, R.L.: Autocorrection in MR imaging: adaptive motion correction without navigator echoes. Radiology 215 (2000) 904–909
9. Blumenthal, J.D., Zijdenbos, Al, Molloy, E., Giedd, J.N.: Motion artifact in magnetic resonance imaging: implications for automated analysis. Neuroimage 16 (2002) 89–92
10. Preboske, G.M., Gunter, J.L., Ward, C.P., Jack Jr., C.R.: Common MRI acquisition nonidealities significantly impact the output of the boundary shift integral method of measuring brain atrophy on serial MRI. Neuroimage 30 (2006) 1196-1202
11. Holdsworth, D., Norley, C., Frayne, R., Steinman, D., Rutt, B.: Characterization of common carotid artery blood-flow waveforms in normal human subjects. Physiological Measurement 20 (1999) 219–240

A simulation framework for diffusion MRI

Matt G Hall* and Daniel C Alexander

Centre for Medical Image Computing (CMIC)
Department of Computer Science
University College London
Gower Street
London WC1E 6BT

Abstract. This note describes a Monte-Carlo simulation framework for diffusing spins that may be used to test and calibrate techniques based on diffusion MRI. This model differs from previously published work in that a mechanism for exchange of spins between "constrained" and "unconstrained" subpopulations of spins is included in the form of a parameter called the membrane permeability. After describing the model in detail, we use it to find optimal parameters of the pulsed-gradient spin-echo sequence. First, we show that it is desirable to have gradient pulses that are as wide and flat as possible, and then show that the b value that maximises the precision of the reconstructed directions is consistent with that estimated using Gaussian test functions.

1 Introduction

Diffusion-weighted MR Imaging is a method of imaging the diffusive motion of water *in vivo* in the direction an applied field gradient. Using a suitable reconstruction technique, diffusion-weighted images may be used to measure the diffusion constant in each voxel. By combining diffusion-weighted imaging sensitised in different directions it is possible to reconstruct the directional dependence of diffusion in scan voxels and thereby infer information about the underlying brain structure by assuming that the diffusion of spins is principally parallel to local microstructure. The standard technique for doing this is diffusion tensor imaging (DTI), which fits a diffusion tensor to six or more diffusion-weighted images on the assumption of Gaussian-distributed water molecule displacements. The principal direction of the diffusion tensor provides an estimate of white matter fibre orientations, which tractography techniques use to reconstruct macroscopic fibre trajectories through the image and then infer brain connectivity. These techniques have emerged as powerful tools for probing pathology [1] and anatomy [2, 3].

Validation and calibration of reconstruction and tractography algorithms often uses simulated data. Most simulations use the Gaussian model of particle displacement to construct synthetic test data. However, the Gaussian model over-simplifies the complex diffusion processes in brain tissue. Here we model brain tissue structure directly and simulate spin diffusion processes within to obtain more realistic synthetic data. We show two simple applications of the new simulation technique to optimisation of parameters of the standard pulse sequence in diffusion MRI.

Section 2 reviews diffusion MRI and diffusion tensor imaging and section 3 describes the new model for simulation and its parameters. In section 4 we describe the experiments performed and present results. Conclusions are in section 5.

2 Background

Diffusion-weighted MRI measures the local properties of water diffusion via a pulsed gradient spin-echo (PGSE) MR scan sequence. Fig-1 illustrates a pulsed gradient spin-echo sequence and the timescales associated with it. An initial 90° pulse aligns spin phases. A spin-position encoding field gradient vector $\mathbf{G}$ of strength $G = |\mathbf{G}|$ is switched on at time t_1, turned off at time $t_2 = t_1 + \delta$, on again at $t_3 = t_1 + \Delta$ and switched off at $t_4 = t_3 + \delta$. The gradient pulse introduces a phase offset to each spin that is linearly proportional to the component of the spin's position in the direction of the applied gradient. A 180° RF pulse between the gradient blocks negates the phase of each spin. Thus, a stationary spin will end up with no phase offset whereas a moving spin will have a residual phase offset proportional to the distance moved between gradient pulses, and the net magnetisation of diffusing spins is attenuated, encoding the diffusion.

The fundamental relationship in diffusion MRI is that the diffusion-weighted measurements are samples of the Fourier Transform of the particle displacement density at wavevector $\mathbf{q} = \gamma\delta\mathbf{G}$ where γ is the gyromagnetic ratio for protons in

*e-mail: m.hall@cs.ucl.ac.uk

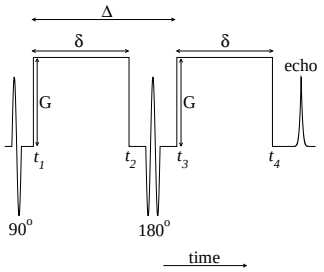


Figure 1. Schematic representation of a pulsed gradient spin-echo (PGSE) MR pulse sequence.

water. This assumes the pulse is negligible so that spins do not move during gradient pulses. This assumption is broken in practice, but the analysis in [4] suggests that this violation may in fact improve estimation of fibre orientations.

In DTI a Gaussian form for spin displacement distribution is assumed, leading to an analytic relationship between measured signal A and spin displacement distribution

$$A(\mathbf{q}) = A(\mathbf{0}) \exp(-t\mathbf{q}^T D\mathbf{q}) = A(\mathbf{0}) \exp(-b\widehat{\mathbf{q}}^T D\widehat{\mathbf{q}}) \tag{1}$$

where $A(\mathbf{0})$ is an unweighted signal ($G = 0$), t is the diffusion time given by Δ, $\mathbf{q} = \gamma\delta G\widehat{\mathbf{q}}$ is the wavevector of the applied field gradient with diffusion sensitisation $b = |\mathbf{q}|^2 t$ s mm^{-2} and D is the diffusion tensor. Taking logs of $A(\mathbf{q})/A(\mathbf{0})$ makes fitting diffusion tensors to measurements a simple linear inversion.

A range of combinations of δ and Δ give the same b. Analysis in [5] shows that taking the pulse width δ as long as possible maximises SNR of fixed b. Increasing b increases directional contrast in diffusion-weighted measurements but also decreases SNR both because attenuation increases and also because higher b value requires a longer echo time. Simulations in [5] use the Gaussian displacement model to show that $b \sim 1000$ s mm^{-2} maximises the concentration of fibre orientation estimates in brain tissue.

3 Methods

We use a Monte-Carlo simulation of spins in a constraining environment to simulate diffusion-weighted measurements. Liu et al [6] have employed a similar technique to produce synthetic data. Their method differs from ours in that spins are restricted to being entirely contained within cylindrical structures, without a freely-diffusing population outside the cylinders and so may tend to over-estimate diffusion anisotropy.

We model the underlying diffusion process measured by a PGSE sequence by considering a population of spins modelled as random walkers in a three dimensional environment. The environment contains a static configuration of barriers which impede the motion of spins. We will refer to this configuration of barriers as the "substrate". The exact configuration of barriers can be selected to reflect a particular system of interest. Here we consider geometries based on a cubic lattice of cubic "cells" of finite size. The lattice consists of an $L \times L \times L$ set of cubic cells with barriers at each cell wall. The edges of each cell have length l.

By removing barriers from opposite ends of cells in lines we can make the spin displacement distribution anisotropic by forming "lanes" of empty space. On a substrate with longitudinal lanes, the motion of spins is restricted in directions perpendicular to lanes, but is free inside lanes along their length (see fig-2). This is representative of diffusion along parallel cylinders, as is found in white matter axon bundles. The barrier configuration in periodic. Spins are initially uniformly distributed across the substrate and then updated according to the following rule (see fig-3):

1. Generate a step vector, $\mathbf{\Delta r}$. Steps are of constant length ℓ and random orientation.
2. Check if step crosses a barrier.
   - If no barrier is crossed, the walker executes the step, walker position $\mathbf{r} \rightarrow \mathbf{r} + \mathbf{\Delta r}$.
   - If the step would take the walker across a barrier, the step is executed with probability p or the walker is elastically reflected by the barrier with probability $1 - p$.
     Walker's position $\mathbf{r} \rightarrow \mathbf{r} + \mathbf{\Delta r}'$ where $\mathbf{\Delta r}'$ is the vector of the reflected step (see fig-3).

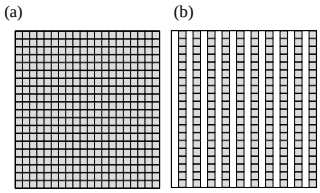


Figure 2. Schematic illustrations of the type of substrates used to generate results. (a) shows an isotropic cellular lattice, and (b) shows a lattice with lanes. Grey cells have barriers, whereas those with barriers removed at opposite ends are coloured white. In three dimensions, layers are interleaved so as to produce square-cylindrical lanes with a checker-board pattern when viewed end-on.

Steps are generated by constructing a randomly oriented vector of fixed length ℓ ($\ell \ll l$) so that spins are free to explore the interior of cells. Steps are amended if a walker is reflected by a barrier. Since ℓ defines the effective diffusion constant for the population, we set it to

$$\ell = \sqrt{\frac{D_{water} \cdot \tau}{T}}, \tag{2}$$

where D_{water} is the free-water self-diffusion constant at body temperature, τ is the duration of the PGSE sequence and T is the number of timesteps in the simulation. Although it would be desirable to generate results from a full molecular-dynamics simulation, unfortunately for systems large enough to simulate a voxel over the duration of a typical PGSE scan this is not computationally feasible, and so we make the Monte-Carlo approximation instead.

A synthetic diffusion-weighted signal is obtained from the simulation by considering the phase of all spins contained in the central region. In common with Liu et al [6] the phase of each spin during the gradient pulses and the final phase offset is proportional to the average position during the pulse. The phase offset of each spin is

$$\Phi_j = \gamma \mathbf{G} \cdot \left(\sum_{t=t_1}^{t_2} \mathbf{x}_j(t) - \sum_{t=t_3}^{t_4} \mathbf{x}_j(t) \right) dt, \tag{3}$$

where $\mathbf{x}_j(t)$ is the position of spin j at time t and $dt = \frac{\tau}{T}$ is the length of timestep.

A noiseless diffusion-weighted signal, S is then $A(\mathbf{G}) = \sum_j e^{i\Phi_j}$. To obtain noisy measurements, real and imaginary Gaussian noise terms with variance σ are added and the modulus of the resulting signal taken following the standard noise model of [7]. We also calculate an unweighted signal $A(\mathbf{0}) = N$, where N is the number of spins in the central region to which noise is added in the same way. We use a single set of spin excursions to generate synthetic signals for many gradient directions.

4 Experiments

We hypothesise that longer gradient pulses lead to more accurate estimates of fibre orientations, as is implied by [4], but not explicitly demonstrated. Furthermore, we hypothesise that the parameter optimisation scheme used by [5] leads to a higher optimal b value for direction concentration when data is synthesised using Monte-Carlo simulation rather than Gaussian test functions.

In all experiments we consider populations of 10000 spins on lattices of $200 \times 200 \times 200$ cells in which spins are

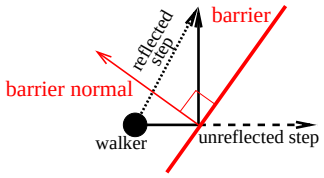


Figure 3. If an update step takes a walker across a barrier it executes the step with probability p or is elastically reflected. The length of the reflected step is such that the sum of the lengths of its components is the same as the length of the unreflected step.

initially distributed uniformly across the lattice and updated 100000 times during the pulse sequence. The cell size is $l = 10^{-5}$m and the barrier permeability is $p = 0$. Substrates have lanes parallel to the y-direction.

We generate noise-free measurements from the simulation and generate 1000 sets of synthetic measurements adding independent noise in each trial. Diffusion tensor elements are then fitted to the signals using the procedure of [8]. By diagonalising the fitted tensors we obtain a set of eigenvectors defining the principal directions of spin displacement. The eigenvector corresponding to the largest eigenvalue gives the fibre orientation estimate.

We compute the precision of estimates of principal diffusion direction over multiple trials using the mean dyadic tensor $\mathbf{Y}$ [5] with elements

$$Y_{ij} = \frac{1}{Z} \sum_{k=1}^{Z} e_i^{(k)} e_j^{(k)} \tag{4}$$

where $e_i^{(k)}$ is the ith component of the principle eigenvector $\mathbf{e}^{(k)}$ of the kth diffusion tensor $D^{(k)}$. The largest eigenvalue κ of $\mathbf{Y}$ is a statistic of direction concentration [5].

4.1 Application to gradient block duration

In order to explore the effect of the duration of the gradient pulse in the PGSE sequence on the direction concentration of fibre orientation estimates we vary G and δ such that the shape of gradient pulses changes shape from being as brief as possible, so as to conform to the assumptions of the Fourier relationship between spin displacement distribution and diffusion-weighted measurements, to being as long and flat as possible in keeping with [4] and [5]. We keep the area under the pulses $G\delta$ constant. The diffusion time Δ is also constant. This issue has remained an unanswered question in diffusion MRI, and this is the first time it has been directly addressed.

We find that the direction concentration increases as block duration in increased (see fig-4(a)). Indicating that long gradient blocks are desirable to maximise the precision of fibre orientation estimates from diffusion tensor measurements.

4.2 Application to parameter optimisation

Using the scheme of Alexander and Barker [5] we generate values of δ and Δ that optimise SNR for various b values, given the standard physical parameter values for white matter (see [5]). Values of δ, Δ and SNR used to generate results are shown in table-1. Parameters were chosen such that SNR=16 when $b = 1000$s mm^{-2}. Data are acquired in 6 different directions.

Fig-4(b) shows the directional concentration κ against the b value for each scan. We see that the direction concentration, κ peaks around $b = 1600$ s mm^{-2}. This is slightly higher than the value obtained by Alexander & Barker [5], who observed a peak at $b = 1040$s mm^{-2} [5].

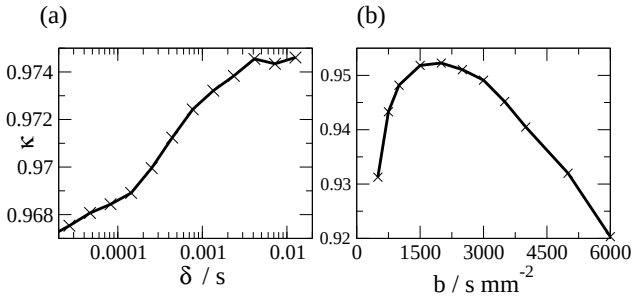


Figure 4. (a) Dependence of direction concentration κ on gradient block shape and (b) Results of applying the simulation to parameter optimisation. We see that the direction concentration peaks at a similar b value to that seen by [5] but the direction concentrations are generally lower in the results from simulation.

b / s mm^{-2}	δ / s	Δ / s	SNR
500	0.0231	0.0271	18.74510
750	0.0266	0.0306	17.15686
1000	0.0294	0.0334	16.00000
1500	0.0338	0.0378	14.33333
2000	0.0374	0.0414	13.11765
2500	0.0403	0.0443	12.17647
3000	0.0429	0.0470	11.41176
3500	0.0453	0.0493	10.76471
4000	0.0474	0.0514	10.19608
5000	0.0512	0.0552	9.294118
6000	0.0544	0.0584	8.549020

Table 1. Parameters used to generate data for parameter optimisation

5 Discussion & Conclusions

We have described a Monte-Carlo model of diffusing spins on a substrate of permeable barriers which restrict their motion. We have seen that the geometry of barriers can be chosen so that a simulation models unrestricted free diffusion, highly restricted directional diffusion or systems in which populations of restricted and freely diffusing spins are present and that using a simulation of spins executing Brownian motion on such a substrate can be of use in generating synthetic data to optimise PGSE scan parameters.

We have used the model to address the issue of finite gradient pulse width in PGSE acquisition sequence and found that longer gradient pulses lead to higher direction concentration in fibre orientation estimate, and furthermore that the optimal b value for maximising direction concentration is similar to that obtained using Gaussian test functions.

The model's utility goes beyond the simple examples addressed in this article. By defining substrates with known geometries, and by making geometries more realistic than those used here, it could readily be applied to calibration of tractography algorithms (e.g. [9]), and on a larger scale also to their validation by using larger substrates representing populations of fibres that are, for example, parallel, diverging or converging as well as more complex situations such as crossing fibres. Another possibility is to construct data from several adjacent voxels that represent some larger structure to be reconstructed by a tractography algorithm. Here a multi-voxel substrate could be used as a ground-truth to compare to reconstructed structure. The model is implemented as part of the Camino Diffusion MRI toolkit [10], and may be downloaded from [11].

Acknowledgements

x MGH is supported by EPSRC grant number GR/T22858/01. The authors gratefully acknowledge the input of Prof Henrik Jensen of Imperial College London Dept of Mathematics for important input regarding the structure of the model.

References

1. J. Kucharczyk, J. Mintorovitch, H. Asgari et al. "Diffusion/perfusion MR imaging of acute cerebral ischemia." *Magn Reson Med* **19**, pp. 311–315, 1991.
2. S. Mori, B. J. Crain, V. P. Chacko et al. "Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging." *Annals of Neurology* **45**, pp. 765–269, 1999.
3. M. Catani, R. J. Howard, S. Pajevic et al. "Virtual *in vivo* interactive dissection of white matter fasciculi in the human brain." *NeuroImage* **17**, pp. 77–94, 2002.
4. P. Mitra & B. Halperin. "Effects of finite gradient-pulse widths in pulsed-field gradient diffusion measurements." *J. Magn. Reson. Ser A* **113**, pp. 94–101, 1995.
5. D. C. Alexander & G. J. Barker. "Optimal imaging parameters for fiber-orientation extimation in diffusion MRI." *NeuroImage* **27**, pp. 357–367, 2005.
6. C. Liu, R. Bammer, B. Acar et al. "Characterizing non-gaussian diffusion by using generalized diffusion tensors." *Magnetic Resonance in Medicine* **51**, pp. 924–937, 2004.
7. J. Sijbers, A. J. denDecker, J. VanAudekirke et al. "Estimation of the noise magnitude in MR images." *Magn Reson Med* **16**, pp. 87–90, 1998.
8. P. J. Basser, J. Matiello & D. LeBihan. "MR diffusion tensor spectroscopy and imaging." *Biophysical Journal* **66**, pp. 259–67, 1994.
9. G. J. M. Parker & D. C. Alexander. "Probabilistic anatomic connectivity derived from the microscopic persistent angular structure of cerebral tissue." *Philosophical Trans Roy Soc B* **360**, pp. 893–902, 2005.
10. P. A. Cook, Y. Bai, S. Nedjati-Gilani et al. "Camino: Open-source diffusion-MRI reconstruction and processing.", 2006.
11. "http://www.cs.ucl.ac.uk/research/medic/camino."

A General Framework for Multiple-Fibre PICo Tractography

K. K. Seunarine[a*], P. A. Cook[a], K. Embleton[b], G. J. M. Parker[b], D. C. Alexander[a]

[a]Centre for Medical Image Computing, University College London, Gower Street, London, WC1E 6BT
[b]Department of Imaging Science and Biomedical Engineering, University of Manchester, Oxford Road, Manchester, M13 9PT

Abstract. Probabilistic Index of Connectivity (PICo) tractography calculates the probability of connection of a voxel in the brain to specified seed voxels. The basic method uses the diffusion tensor to estimate the orientation of the white-matter fibres in the voxel. However, the diffusion tensor can only model a single fibre-orientation and fails at fibre-crossings. Here, we describe a general extension to the PICo method which exploits the information from multi-fibre reconstruction techniques such as Persistent Angular Structure (PAS) MRI or Q-Ball imaging, which resolve orientations of multiple fibres in each voxel. We show results for PICo tractography using PASMRI on human brain data and compare the results to those of the original algorithm.

1 Introduction

Diffusion MRI probes the microstructure of the brain by measuring the diffusion of water molecules. In particular, it is used to examine the fibrous white-matter that connects the different regions of the brain. The technique has several applications; it can be used to detect white-matter diseases [1], generate atlases of connectivity of the brain [2] and in neurosurgical planning and evaluation [3, 4]. The standard technique in diffusion MRI is Diffusion Tensor Imaging (DTI), which fits the diffusion tensor to diffusion-weighted measurements. The principal direction of the diffusion tensor provides an estimate of the dominant fibre orientation in white-matter regions. Tractography algorithms use these fibre-orientation estimates to reconstruct the paths of white-matter fibre-tracts by following estimates from voxel to voxel through the image.

However, DTI fails when voxels contain complex structure such as crossing fibres because it can only compute a single fibre direction in each voxel. To overcome this limitation, various techniques have been developed, such as Persistent Angular Structure (PAS) MRI [5] and Q-Ball imaging [6], which are able to resolve multiple fibre orientations in a single voxel. However, most tractography algorithms still use the basic DTI single-fibre reconstruction and it is not clear how to generalize them to exploit the extra information that multiple fibre reconstructions provide. This paper describes a general framework for probabilistic tractography using multiple fibre orientations. The next section describes background and previous work on which we build. Section 3 develops the framework. In section 4, we describe application of the framework to a human brain data set and compare tractography results from the new framework using PASMRI with the original PICo algorithm using DTI. Finally, in section 5 we present some conclusions and a discussion of further work.

2 Background

DTI fits the diffusion tensor (DT) and uses the principal eigenvector to provide an estimate of the fibre-orientation for a voxel. The drawback of DTI is that it can not resolve configurations such as crossing or kissing fibres and fibre bundles that splay. One alternative is to use a multi-tensor model [7], but this requires prior knowledge of how many fibre populations are present in a given voxel. A variety of techniques, such as Diffusion Spectrum Imaging [8], Q-Ball, Spherical Deconvolution [9] and PASMRI aim to reconstruct the distribution of fibre orientations in each voxel without prior knowledge of the number of fibre directions. Each of these algorithms returns a function on the sphere which reflects the distribution of fibre orientations. However, the precise relationship between these functions and the true fibre orientation distributions is not clear. In particular, the shapes of the spherical distributions that the methods return are very different and depend on parameter settings in the reconstruction algorithms; see [10] for examples. One feature of the distributions that is reliable, however, is the peak directions, which provide good estimates of the dominant fibre orientations [10].

Tractography uses the fibre-orientation estimates generated during reconstruction to find the paths of white-matter tracts through the brain. Starting from a seed point, a path (or streamline) is propagated from voxel to voxel using the fibre orientation estimates as the direction of each step. Noise in the diffusion-weighted measurements causes uncertainty in the estimates of the fibre orientations, so the streamline may not reflect the true path of the white-matter.

*Email: k.seunarine@cs.ucl.ac.uk

Probabilistic Index of Connectivity (PICo) tractography [11] generates a map of probability of connection of each voxel to a specified seed voxel. Starting from a seed point, N streamlines are generated. For each streamline, the fibre-orientation estimates are randomly perturbed according to a probability density function (PDF) of the error, which is indexed from a lookup table. The probability Φ of a voxel $\mathbf{v}$ being connected to the seed voxel is then

$$\Phi(\mathbf{v}) = \lim_{N\to\infty} \Phi(\mathbf{v}, N) \approx \frac{\chi(\mathbf{v}, N)}{N}, \tag{1}$$

where $\Phi(\mathbf{v}, N)$ is the probability of connection given N streamlines and $\chi(\mathbf{v}, N)$ is the number of occasions at which $\mathbf{v}$ is crossed by a streamline. The algorithm predicts the uncertainty of each fibre orientation estimate from the fractional anisotropy (FA) [12] of the DT. High FA suggests high confidence in the estimate and low FA suggests low confidence or high uncertainty. More precisely, a calibration phase uses synthetic data to measure the correlation between FA and the uncertainty in fibre orientation caused by image noise.

Several extensions to PICo tractography have been suggested. Cook et al [13] describe a similar procedure that uses more standard spherical distributions such as the Watson and Bingham distributions [14] instead of the Gaussian to model fibre-orientation uncertainty. They demonstrate the procedure using both synthetic data and for PICo tractography on brain data and show that the Watson PDF provides a better estimate of uncertainty than the Gaussian PDF.

Parker and Alexander [15] show that the PICo algorithm generalizes easily to use multi-tensor models, which can improve tracking through fibre-crossings. However, the combined framework suffers from the limitations of the multi-tensor models such as model-fitting instability and the need to prespecify the number of fibres per voxel. More recently, Parker and Alexander [16] extended PICo tractography for use with multi-fibre reconstruction algorithms such as PASMRI. To exploit these techniques, they use peak directions of the reconstructed fibre-orientation distributions, of which there may be several in each voxel, as estimates of fibre orientations. In place of the FA of the diffusion tensor, they use the shape of the fibre orientation distribution peak to predict the uncertainty of the estimate it provides. Broad peaks have a high uncertainty, sharp peaks have a low uncertainty. Specifically, they compute the Hessian, or second partial derivative matrix, at each peak and use its trace as a measure of mean curvature and predictor of uncertainty. The lookup table of variances of Gaussian uncertainty PDFs is indexed by the log of the trace of the Hessian of each peak and is constructed, as usual, using simulations.

A weakness of the approach in [16] is that it uses the Gaussian PDF, which does not account for anisotropy of the fibre-orientation-estimate uncertainty and the contours of the PDFs are all circular. Here, following Cook et al [13], we generalize the method to use a standard distribution on the sphere, the Bingham distribution, which can model more general elliptical clusters of fibre-orientations.

3 Methods

The Bingham distribution is

$$p(\mathbf{x}) = \frac{1}{M_2(\kappa_1, \kappa_2)} \exp[\kappa_1(\boldsymbol{\mu_1} \cdot \mathbf{x})^2 + \kappa_2(\boldsymbol{\mu_2} \cdot \mathbf{x})^2], \tag{2}$$

where M is the confluent hypergeometric function of the first kind [14], the parameters $\kappa_1 \leq \kappa_2 \leq 0$ define the concentration of the distribution. The mean of the distribution is $\boldsymbol{\mu_1} \times \boldsymbol{\mu_2}$, where $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ are the principal axes of the elliptical distribution normal to the mean. The eccentricity of the contours of the probability density is defined by $\kappa_2\kappa_1^{-1}$. When $\kappa_2\kappa_1^{-1} \approx 1$ the distribution is circular with a peak normal to the mean. As $\kappa_2\kappa_1^{-1} \to 0$ the contours become increasingly eccentric. If $\kappa_1 = \kappa_2 = 0$ the distribution is uniform. Kent [17] shows how to calculate the maximum likelihood estimates of the parameters κ_1, κ_2, $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$.

The new framwork we propose uses the Hessian matrix at each fibre-orientation distribution peak to predict the parameters of a Bingham model of uncertainty in the fibre-orientation estimate that peak provides. The eigenvectors of the Hessian are the semi-major and semi-minor axes of an elliptical model of the peak cross section. We use the Hessian eigenvectors as $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ in the Bingham model directly. To predict the concentration parameters κ_1 and κ_2, we construct a lookup table mapping the Hessian eigenvalues to values of κ_1 and κ_2.

As in previous PICo implementations, we construct the lookup table using simulations on functions with known peak directions. Specifically, for a large number of noisy trials we reconstruct fibre-orientation estimates and associated Hessian matrices using the multiple-fibre reconstruction of choice. For each peak in each function, we compute the rotation that rotates the peak direction to the z-axis and the major eigenvector of the peak Hessian to the x-axis. We

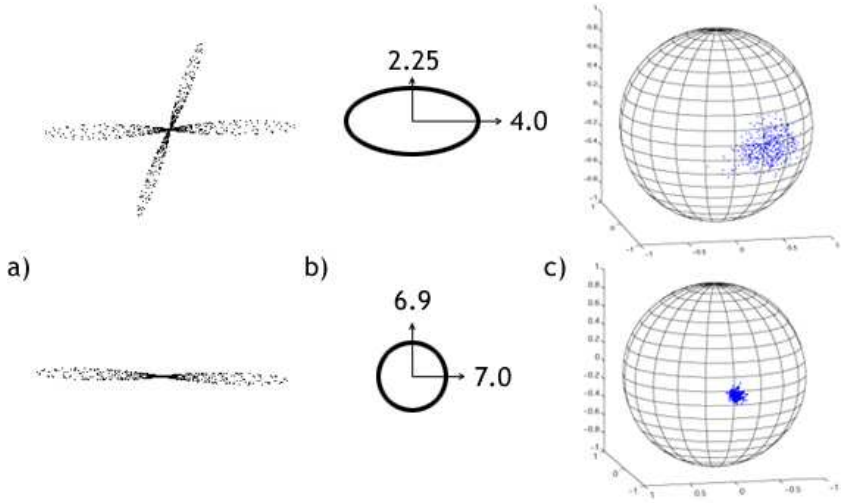


Figure 1. a) the fibre orientation distribution, b) the shape of the cross-section at the peak, with the principal eigenvectors and the log of the eigenvalues shown, and c) samples from the Bingham distribution using parameters indexed from the lookup table.

rotate the true direction that the peak provides an estimate for by that same rotation. Thus we obtain a population of rotated true fibre directions each with an associated pair of Hessian eigenvalues. We construct a 2D histogram indexed by the log of the two Hessian eigenvalues to obtain a population of rotated true fibre directions in each bin. We fit the parameters of the Bingham distribution in each bin containing 50 or more samples. Finally, we fit a quadratic surface to the log of each κ_1 and κ_2 as a function of the Hessian eigenvalues.

Figure 1 illustrates the mapping from fibre-orientation distribution (PAS in this example) to fibre direction PDF. Figure 1(a) shows two example PAS reconstructions from test functions used in the PICo calibration. Figure 1(b) shows the shape of the cross section of a peak of each PAS function. Figure 1(c) shows samples from the Bingham distribution predicted by the lookup table for each PAS peak.

4 Experiments and results

In this section, we illustrate the new framework using PASMRI reconstruction in a normal human brain data set and compare results with those using DTI. We emphasize however, that the general framework we outline above can exploit other multiple-fibre reconstructions using exactly the same procedure.

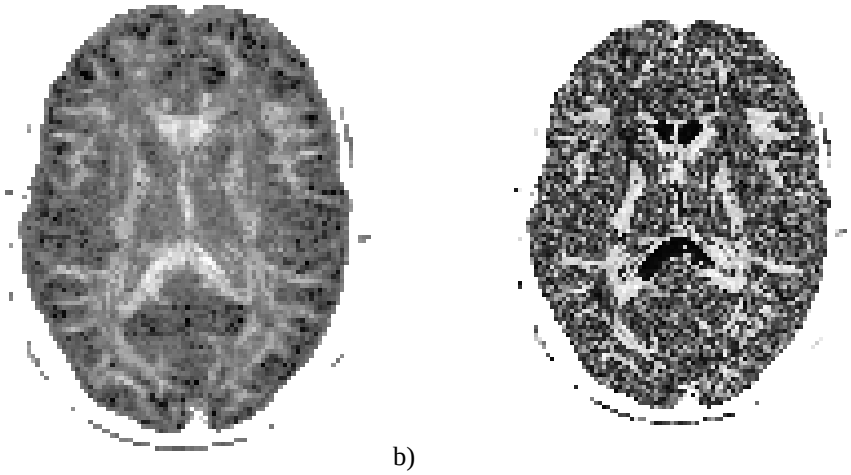
The synthetic data used in the configuration of the lookup table contain either one or two fibre orientations per voxel. The datasets containing two fibre orientations per voxel were generated using variations of the test function

$$p(\mathbf{x}) = aG(\mathbf{x}; \mathbf{D}_1, t) + (1-a)G(\mathbf{x}, \mathbf{R}_\theta^T \mathbf{D}_2 \mathbf{R}_\theta, t), \tag{3}$$

where a is a mixing parameter, $G(\mathbf{x}; \mathbf{D}, t)$ is a zero mean Gaussian with covariance $2t\mathbf{D}$, $\mathbf{D}_1 = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_2)$, $\mathbf{D}_2 = \mathrm{diag}(\lambda_2, \lambda_1, \lambda_2)$ and $\mathbf{R}_\theta$ is a rotation by θ about the z-axis. We use combinations of $\lambda_2 \in \{1, 3, 4, 5\} \times 10^{-10}\,\mathrm{m^2\,s^{-1}}$, $a \in \{0.5, 0.6\}$ and $\theta \in \{0, 22.5^o, 45^o\}$. For datasets containing a single fibre orientation per voxel, the test function is $p(\mathbf{x}) = G(\mathbf{x}; \mathbf{D}_1, t)$, where $\lambda_2 \in \{1, 3, 4, 5\} \times 10^{-10}\,\mathrm{m^2\,s^{-1}}$. In all of the datasets $\lambda_1 + 2\lambda_2 = 2.1 \times 10^{-9}\,\mathrm{m^2\,s^{-1}}$, which is approximately the value expected in brain data. Datasets were generated for 100 random rotations of each setting of the test function, and each dataset contains 10 sets of measurements. The data is synthesized by sampling the Fourier transform of p at each wavenumber in a spherical acquisition scheme with 61 gradient directions and $b = 1200\,\mathrm{s\,mm^{-2}}$ according to the standard Fourier relationship between the particle displacement density p and the diffusion weighted measurement [18]. Noise is added to the measurements as random complex numbers with independent real and imaginary parts drawn from the normal distribution $N(0, \alpha^2)$, where $\alpha = F(0)/S$, F is the Fourier transform of p at each wavenumber, S is the signal to noise ratio (SNR) at $b = 0$. We then take the modulus to get the synthetic measurement. The data generated uses $S = 20$, which matches the SNR of the brain data used for tractography.

Brain data was acquired along 61 gradient directions and a single acquisition with $b = 0$. Diffusion-weighted imaging

was performed using a SE echo-planar imaging sequence with TE=54ms, TR=11884ms and $b = 1200\,\mathrm{s\,mm^{-2}}$. 60 axial slices were imaged. A 112×112 acquisition matrix was used, which was interpolated to 128×128. The voxel size is $1.875 \times 1.875 \times 2.1\mathrm{mm}^3$. The total imaging time was approximately 28 minutes.



a) b)

Figure 2. Maps of a) the normalized principal eigenvalue of the Hessian at the largest PAS peak and b) the Hessian anisotropy for an axial slice of the human brain data.

Figure 2(a) maps $\log(-h_1/\bar{\Phi})$, where h_1 is the most negative eigenvalue of the Hessian matrix at the largest peak of the PAS function Φ and $\bar{\Phi}$ is the mean of Φ. The quantity $\log(-h_1/\bar{\Phi})$ is the predictor of κ_1 and high intensity in figure 2(a) corresponds to high confidence. We see high confidence in the major white matter bundles such as the corpus callosum, as we expect, which suggests the prediction is good. Figure 2(b) shows the fractional anisotropy [12] of the Hessian at the largest peak and is brightest in regions where the peak has highly elliptical cross section. The intensity is high in several fibre crossing and partial volume regions, as well as in the cortico-spinal tract, which diverges at the level of this slice as it runs towards the top of the brain. The Hessian anisotropy in the centre of the corpus callosum is low, since the fibres are much more parallel to each other.

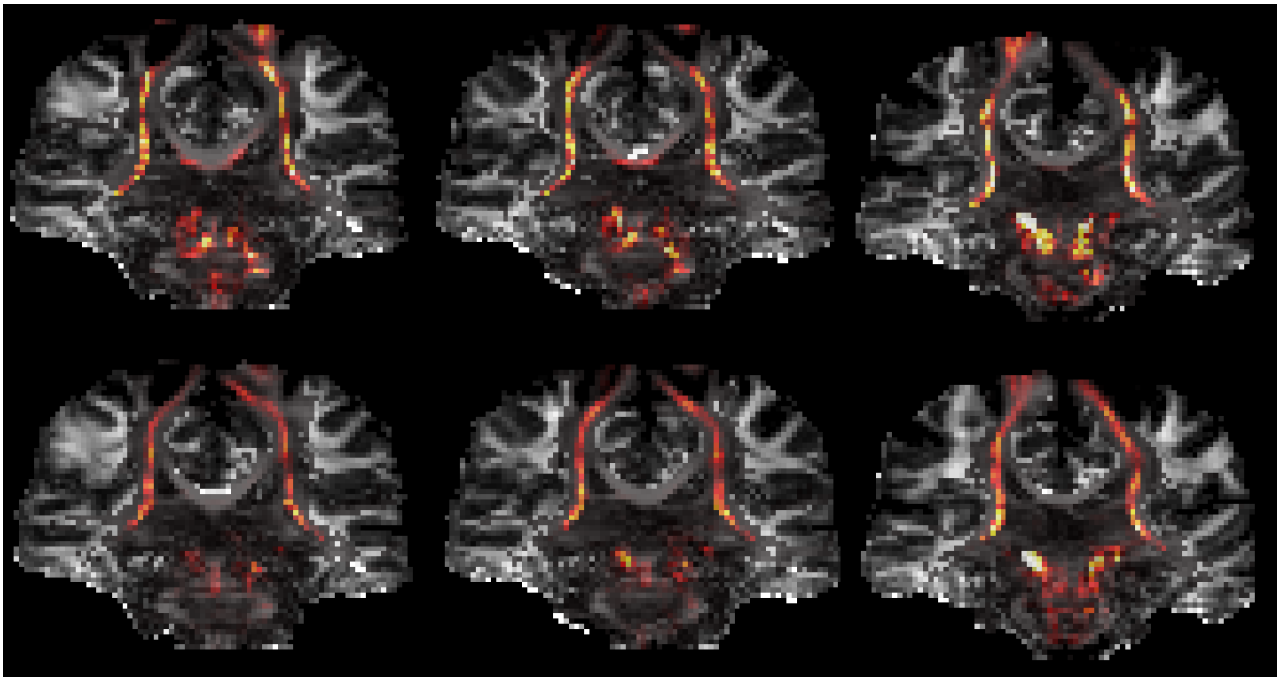


Figure 3. PICo connection probabilities overlaid on the fractional anisotropy map of the subject. The top row is the result of tracking in DTI data, the bottom row is the result of tracking in PAS data.

We run PICO tractography on the human brain data using new multiple-fibre PICo algorithm in conjunction with

PASMRI reconstruction and the original algorithm using DTI. The tractography process was started from a manually defined region of interest at the base of the cortico-spinal tract. The dominant pathway from the seed region runs inferior-superior into the corona radiata, where the cortico-spinal fibres cross lateral fibres projecting from the corpus callosum. Figure 3 shows the results of tractography using the multi-fibre algorithm with PASMRI. The DTI PICo maps show significant connection probabilities in the corpus callosum, which are false-positive connections arising from fibre crossings at the ends of the corpus callosum where it meets the cortico-spinal tract. The DTI PICo maps show the cortico-spinal tract connecting to the motor regions in the parietal cortex [19, Figure 764]. The PAS PICo results show very low connection probabilities in the corpus callosum, which shows better handling of fibre crossings leading to reduced false positive connections and overall more accurate reconstruction of the fibre pathways.

5 Discussion and Conclusions

We have introduced a new method for exploiting the information from multiple-fibre reconstruction techniques such as PASMRI. The technique uses the shape of the peaks of the fibre orientation distributions to predict the uncertainty in the fibre-orientation estimates. We demonstrate the method using the PASMRI reconstruction algorithm, but note that it can be used with any multiple-fibre reconstruction algorithm that outputs a function on the sphere representing the fibre orientation distribution. The methods presented here have been implemented in the Camino [20] open-source diffusion MRI processing toolkit.

Acknowledgements

This work is funded by the EPSRC grant GR/T22858/01.

References

1. Q. Dong, R. C. Welsh, T. L. Chenvert et al. "Clinical applications of diffusion tensor imaging." *Journal of Magnetic Resonance Imaging* **66**, pp. 259–267, 1994.
2. T. E. J. Behrens, H. Johansen-Berg, W. M. Woolrich et al. "Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging." *Nature Neuroscience* **6**, pp. 750–757, 2003.
3. C. A. Clark, T. R. Barrick, M. M. Murphy et al. "White matter fiber tracking in patients with space-occupying lesions of the brain: a new technique for neurosurgical planning?" *NeuroImage* **20**, pp. 1601–1608, 2003.
4. M. Lazar, P. Thottakara, A. S. Field et al. "A white matter tractography study of white matter reorganization after surgical resection of brain neoplasms." In *12th Annual Meeting of the ISMRM*, p. 1259. Kyoto, 2004.
5. K. M. Jansons & D. C. Alexander. "Persistent angular structure: new insights into diffusion MRI." *Inverse Problems* **19**, pp. 1031–1046, 2003.
6. D. S. Tuch. "Q-Ball Imaging." *Magnetic Resonance in Medicine* **52**, pp. 1358–1372, 2004.
7. D. S. Tuch, T. G. Reese, M. R. Wiegell et al. "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity." *Magnetic Resonance in Medicine* **48**, pp. 577–582, 2002.
8. V. J. Weeden, T. G. Reese, D. S. Tuch et al. "Mapping fibre orientation spectra in cerebral white matter with fourier-transform diffusion MRI." In *7th Annual Meeting of the ISMRM*, pp. 259–267. Berkley, USA:ISMRM, Philadelphia, 1999.
9. J. D. Tournier, F. Calamante, D. G. Gadian et al. "Direct estimation of the fibre orientation density function from diffusion-weighted MRI data using spherical deconvolution." *NeuroImage* **23**, pp. 1176–1185, 2004.
10. D. C. Alexander. "Multiple-fibre reconstruction algorithms for diffusion MRI." *Annals of the New York Academy of Sciences* **1046**, pp. 113–133, 2005.
11. G. J. M. Parker, H. A. Haroon & C. A. M. Wheeler-Kingshott. "A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements." *Journal of Magnetic Resonance Imaging* **18**, pp. 242–254, 2003.
12. P. J. Basser & C. Pierpaoli. "Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI." *J. Magn. Reson. B* **111**, pp. 209–219, 1996.
13. P. A. Cook, D. C. Alexander & G. J. M. Parker. "Modelling noise-induced fibre-orientation error in diffusion tensor MRI." In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 332–336. Arlington, 2004.
14. K. V. Mardia & P. E. Jupp. *Directional Statistics*. Wiley, 2000.
15. G. J. M. Parker & D. C. Alexander. "Probabilistic monte-carlo based mapping of cerebral connections utilizing whole-brain crossing fibre information." In *Proc. IPMI*, pp. 684–695. 2003.
16. G. J. M. Parker & D. C. Alexander. "Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue." *Philosophical Transactions of the Royal Society* **360**, pp. 893–902, 2005.
17. J. T. Kent. "Asymptotic expansion for the bingham distribution." *Applied Statistics* **36(2)**, pp. 139–144, 1987.
18. P. T. Callaghen. *Principles of Magnetic Resonance Imaging*. Oxford: Oxford Science Publications, 1991.
19. P. L. Williams, R. Warwick, M. Dyson et al. *Gray's Anatomy, 37th Edition*. Churchill Livingstone, London, 1989.
20. P. A. Cook, Y. Bai, S. Nedjati-Gilani et al. "Camino: open source diffusion-MRI reconstruction and processing." In *Proceedings of the International Society for Magnetic Resonance in Medicine*. Seattle, 2006.

The Hybrid Conceptual Model of the human lung

Spyridon Montesantos[a], John S. Fleming[a], Livia Bolt[a]

[a]School of Medicine, Southampton University Hospitals NHS Trust, Department of Nuclear Medicine, Mail point 29, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK.

Abstract. Models of the airway tree are useful in interpretation and analysis of lung images. The conceptual model of the airway considers each lung divided into ten concentric shells and specifies the volume of each airway generation in each shell. This provides a generic model of the spatial distribution of airways, which can be fitted to any individual lung from knowledge of its outline. The hybrid model presented in this study is a modification of the conceptual model, which has a more realistic volume distribution of airway generations throughout the lung. It also introduces improved data on the individual patient airway by incorporating in vivo CT data of the first few generations. The hybrid model is compared with previous models at both shell and generation levels. The resulting model can be applied in combination with radionuclide imaging to define airway deposition per generation in the lung and should provide improved accuracy of analysis compared to previous models.

1 Introduction

Anatomical models of the lung are a valuable tool in the field of aerosol science. They can be used in aerosol deposition modelling and in analyzing and evaluating experimental deposition data obtained through 3D imaging protocols like SPECT. More importantly, such models help scientists to make steps towards the culmination of selective distribution of aerosol medicine in the lung. These can improve the efficiency of the treatment, decreasing the cost/efficiency ratio and reducing the probability of both overdosing and creating adverse sideffects associated with therapy. The ultimate goal of this line of research is the production of a complete description of the lung airway tree.

The first person to propose an airway structural model that included both airway geometry and connectivity was Weibel [1] in 1963. Since then many efforts of producing an accurate description of the morphology of the human lung have been made. These vary from statistical models (Horsfield et al. [2], Yeh and Schum [3], Phalen et al. [4], ICRP [11]) to complete 3D, fully deterministic branching structures (Martonen et al. [5], Kitaoka et al. [6] and Tawhai et al. [7]). The latter tend to dominate current research done on the field. A different type of approach was attempted by Fleming et al. [8, 9] with the presentation of a conceptual model of the lung for the purpose of it being immediately applied to image analysis. The goal was the development of a methodology for deriving the aerosol deposition per generation for the 3D spatial distribution of deposition measured using SPECT imaging. The model is referred to as conceptual due to the lack of information on the spatial location of the individual airways.

According to this model, the lung volume is subdivided into concentric hemispherical subvolumes called shells, each shell defined at a fixed fractional distance from the hilum to the periphery of the lung. A great advantage of defining the airway this way was that the model could be easily adapted to fit any probable lung shape, which is a fundamental requirement of any model used for SPECT data analysis. However, though the conceptual model has been demonstrated to produce satisfactory results when calculating deposition from 3D image data (Fleming et al. [10]), it had certain weaknesses in its original design such as erroneous assumptions on the locations of specific airways or calculations of airway dimension data based on specific datasets only.

Though most of these weaknesses were addressed during the revision and validation of the conceptual model [9], certain model issues remained unsolved. This study has the purpose of updating the conceptual model in an effort to eliminate most of those issues. Two major improvements were attempted. The first is the substitution of the generic conceptual model with individual anatomical data for the first few airway generations, where this is possible (Sauret et al. [12]). The new model will be referred to as hybrid due to the addition of these real data. The second is the creation of a continuous distribution of the airways which will be more realistic than the discrete distribution in shells assumed in the current conceptual model.

2 Methods

Hybrid conceptual model

The main idea underlying the creation of the conceptual model is the division of the lung into a set of discrete, concentric subspaces, which will be referred to as shells. MRI images of one human subject provided the lung outline. These were manually aligned to the CT images of the same subject, which exhibited the first 6 generations

(0-5) of the airway tree. The lung volume of the CT images is considered to be composed of cubic voxels with linear dimensions of 2.6 mm. The CT data were then incorporated into the airway model. After the alignment, computerized image analysis provided the exact location of the hila, which are considered to be at the points where the main bronchi infiltrate the lungs. The program used for image analysis and algorithm implementation throughout this study is MATLAB version 6.5. Lung airway volumes were also normalised to a standard lung volume of 3300ml Functional Residual Capacity (FRC) to enable comparison with the literature.

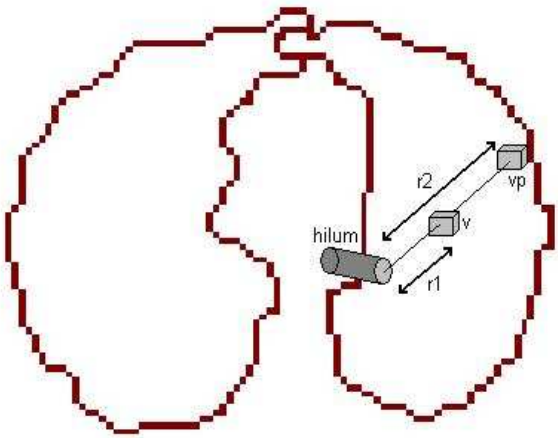


Figure 1. Spherical transformation schematic. v indicates the voxel in question and vp the corresponding voxel in the periphery of the lung.

The shell-like structure in the lung was created using spherical transformation. According to this method, each voxel in the lung will be uniquely mapped to a voxel in a hemisphere which has the same volume as the lung. The mapping will occur with respect to the fractional distance of the voxel from the hilum, which is regarded as the centre of the hemisphere, to the extrapolated periphery of the lung. This is done by creating a straight line passing through both the hilum and the voxel in question and deciding which periphery voxel this line passes through. The fractional distance (r) is then given by the ratio (voxel distance from hilum) / (corresponding periphery voxel distance from hilum) (r1/r2 in Figure 1).

The ultimate goal of the conceptual model is to assign the volume of each generation, j, in each shell, i, Vij. The step immediately after the shell analysis described above is to calculate the number of acini in each shell. For simplicity's sake, only the first 14 generations of lung airways will be regarded as conducting airways while generations 15-23 will be considered as acini (units of alveolated airways). Shell 1 contains no acini. Typical conducting airway paths to acini in shells 2-10 are then identified with the help of morphometric data obtained from previous studies (mainly [12]). Vij could then be evaluated using information on the actual length and diameters. In the hybrid model, calculations of Vij for the first 6 generations are not required since they are provided by CT data.

The original conceptual model assumes an even distribution of each Vij throughout a shell. The volume of every generation contained in every voxel in a single shell was found by dividing the Vij by the total volume of the shell, i.e. with the total number of shell voxels

$$VVij = \frac{Vij}{Ni} \qquad (1),$$

where Ni is the number of voxels in shell i. This resulted in abrupt and, therefore, abnormal changes in air distribution at shell boundaries. To obviate this, an interpolation scheme was devised, using the fractional distances (r) as the basis for all subsequent computations. According to this method, the only voxel values calculated with the help of equation (1) were the voxels at the middle of the shells, i.e. with half values of fractional distances (0.05, 0.15 etc.). After that, a graph is produced for every generation, matching these unique voxel volumes to certain fractional distances. In this sense, the volume of a voxel is considered to be a function of fractional distance (VVj(r)) during the interpolation process. Such graphs, showing the linear implementation of the interpolation algorithm, can be observed in Figure 2 (d) and (e).

Some important assumptions have to be pointed out here. First, we consider all generations to have zero volume at zero fractional distance, except for generation 1 which will have maximum volume. Also, the second half of shell 10 does not contain conducting airways but only acini, which are in fact the sum of generations 15-23. The acini volumes at the edge of the lung (fractional distance 1) are calculated using the following equation

$$VV_{15}(1)=[VV_{15}(0.65)+VV_{15}(0.75)+VV_{15}(0.85)+VV_{15}(0.95)]/4 \quad (2),$$

where VV_{15} is the voxel volume of the acini, which is basically the mean value of the previous four calculated points. The fractional distance of every lung voxel is then compared to the equivalent graph to extract the volume of the specific voxel. This volume value is then saved in a new 3D matrix at the same coordinates as the voxel in question.

Thus the lung is filled with the interpolated generation volumes. A 3D matrix of voxel volumes VVkj is defined for each generation j where k is the representative voxel. Any voxel already identified as containing airways of generation 0-5 is set to zero volume for all other generations. Two examples of this process can be seen in Figure 2(a) and (b). These show the distribution of air volume in a transverse slice near the centre of the lung for the acini ('generation 15') and generation 8. The empty spaces at fractional distances where there would be expected to be some airway for that generation denote positions occupied by the known physical airways which are given by the in vivo data. Figure 2(c) shows a combination of all the known airway generations (0-5) for a single slice.

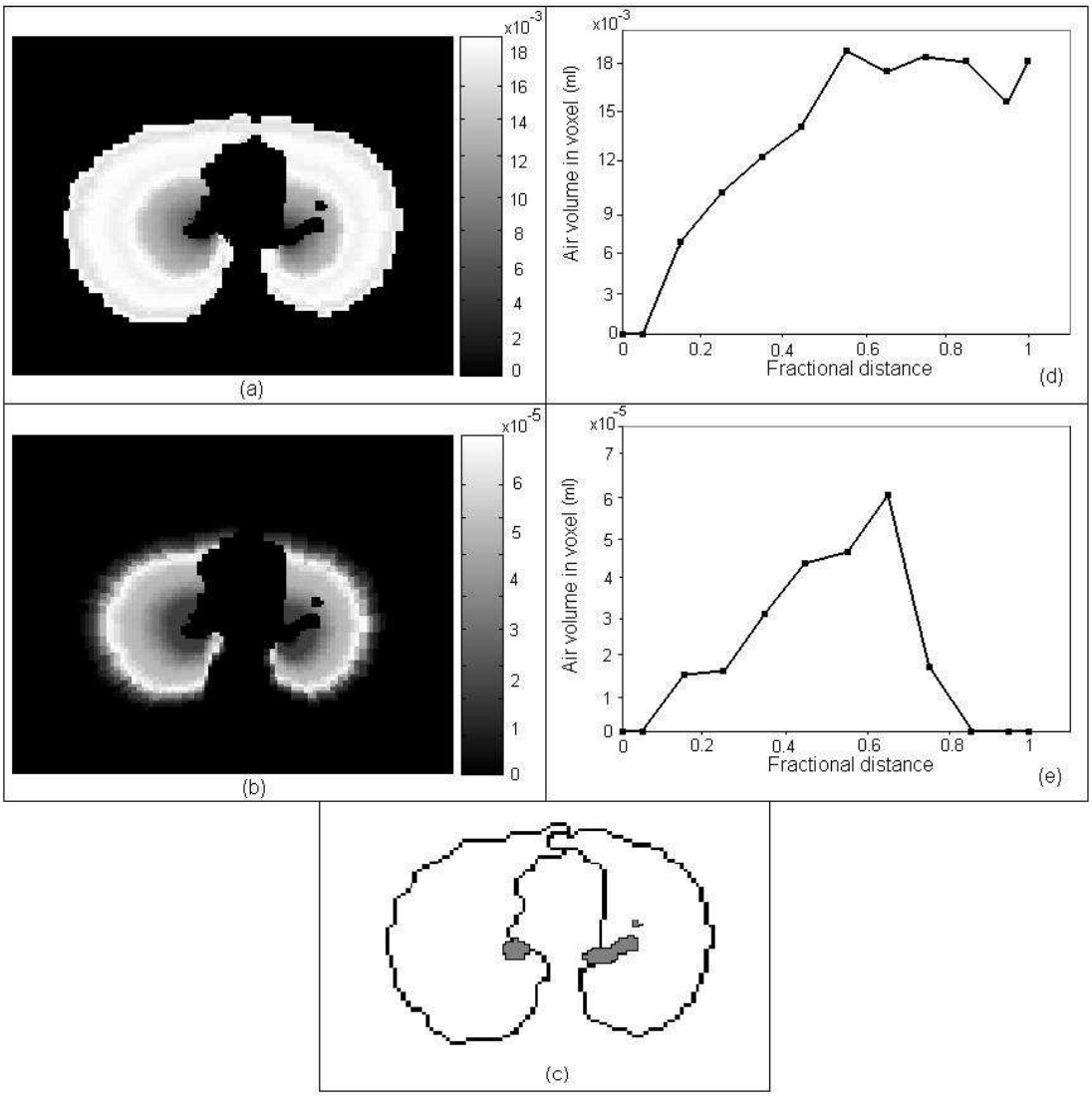


Figure 2. Single slice representation of the improved model in the acini (a) and generations 8 (b), along with the corresponding linear graphs that were used to perform the interpolation (d), (e). Part (c) represents the 6 known airway generations for the same transverse slice. Every color represents different volume values that change with the fractional distance according to the rules illustrated at the corresponding graphs.

Lastly, for comparison purposes, some data on airway geometry derived from spiral CT images of a replica lung cast obtained by Sauret et al. [12] where utilized. The method used was the following: The length, diameter and volume of each airway were determined and images of the airways of each generation up to generation 8 were formed. Then, a 3D shell analysis on these images was performed to determine the volume of each generation within each shell, defining the outer boundary of the lung as the cast envelope dilated by 5mm to allow for the alveolated airways missing from the cast. Finally, the new data were normalized to a standard FRC of 3300 ml.

The variation of the mean fractional distance and volume with airway generation are shown in Figures 3 and 4 respectively. Figure 3 also shows the mean fractional distance for the revised conceptual model and actual CT image data obtained from a lung cast using the method described previously. Figure 4 includes data on volume from previous literature models of the airway tree for comparison purposes. All data are normalised to the same FRC.

3 Discussion

Generic models of the lung are useful in providing estimates of airway anatomy. However, it is widely accepted that there is considerable individual to individual variation in lung morphology. Therefore, there are inevitable limitations on the accuracy of generic models. This paper has demonstrated a method of adopting an existing generic model to a specific individual using measurement data on the actual airway structure of the first few generations.

When the mean shell position of each generation is compared between the hybrid, the revised conceptual model and the CT cast data, the only important discrepancies exist around generation 1 (Figure 3). Other than that, the data of all methods follow similar patterns and the values of mean fractional distance are very close to each other. The cast data provides information on airway location up to generation 9. The fact that the mean shell position of each generation predicted by the hybrid model is in close agreement with the actual locations for the cast beyond generation 5 provides evidence of the validity of the model. Therefore, the model can be safely employed in applications demanding accuracy in the calculation of airway depth in the lung. The large disparities in the first generation can be attributed to errors introduced during the determination of the hilum position in the revised conceptual model, the most important of which is the incorrect assumption that the hilum occurs at the bifurcation of the main bronchi.

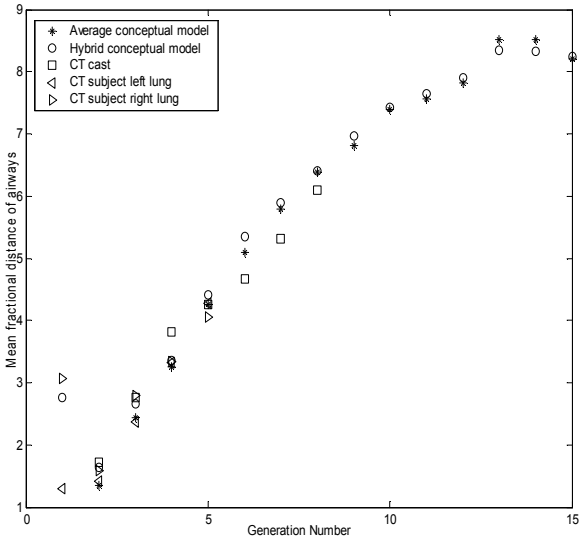


Figure 3. Mean fractional distance of airways vs. Generation number

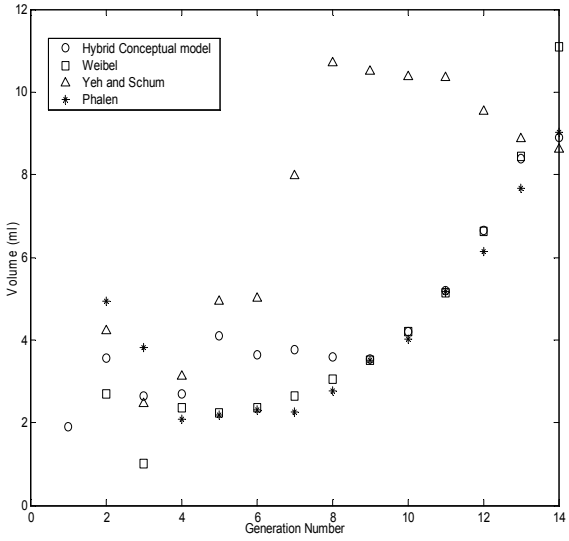


Figure 4. Airway volume in ml vs. Generation number

The comparison of values for volume per generation of the conducting airways for the hybrid model and the other three literature models shows that only the hybrid model can provide information on the amount of generation 1 volume that lies within the lung space. This constitutes an important advantage of the hybrid model as the original version either over [8] or underestimated [9] generation 1. In addition, it is shown in Figure 4 that the hybrid model tracks the Phalen and the Weibel models in high numbered generations of the conducting airways but the in vivo data (generations 0-5) are closer to the Yeh and Schum values. However, though the comparison of values for volume per airway generation for the existing lung models illustrates considerable variation, a significant part of this variability comes from the Yeh and Schum model, which displays differences in the order of several ml between generations 6-14. It is difficult to speculate whether this represents typical inter-subject variability or not. This question can be addressed in the future by carrying out CT imaging of the lung in different subjects. These measurements will give data on this inter-subject variability of volume up to generations 8-9, which includes some of the generations where the variability seems to be the highest.

Generally speaking, the hybrid model has two major advantages over its predecessor. The first one is that it uses real anatomical data provided by CT scans to customize it according to the anatomy of each subject. The second major

advantage of the hybrid model is the improvement of the airway volume distribution throughout the part of the lung where no real anatomical data exist. This addition gives the model a more lifelike quality, increasing thus its usability in clinical applications and its comparability to other, mostly deterministic, models.

In relative terms, the generation volume of the hybrid conceptual model is very similar to the average model presented in [8] and, indeed, with the ICRP model of 1994 [11]. However, the improved modelling of generations 0-5 should lead to reduced errors when using it to estimate deposition per generation. Still, the hybrid model provides data in a different form than the average conceptual model. It describes each generation by a 3D distribution of volume rather than of discrete volume in each shell. This will require new algorithms to be developed to derive deposition per generation values.

4 Conclusion and future work

The next step in this line of research, as far as the hybrid model is concerned, is the validation of the model's results with the help of simulated radionuclide imaging. This has to be done in order for the hybrid model's practicality to be completely established with respect to aerosol deposition in the lung and for the minimization of the errors suggested in the previous paragraph. Another goal that should be pursued is the enhancement of the accuracy of the model by trying to incorporate more generations of real lung airways to it. Finally, further work should include efforts for the better localization of the interpolated airway generations, in an attempt to provide the model with even more lifelike results.

In conclusion, this study introduced a different approach in the implementation of the conceptual model, improving the airway distribution between different shells and introducing in vivo data in the realization of the model. This hybrid model can be used instead of the revised conceptual model and in conjunction with radionuclide imaging to give good measurements of depositions per airway generation in the lung. As CT scanning techniques advance, it will be possible to integrate even more in vivo data in the model, thus improving its performance and reducing the discrepancies involved in its application.

5 References

1. Weibel, E.R. 1963. *Morphometry of the Human lung*. New York, Academic Press.
2. Horsfield, K., G. Dart, D.E. Olson, G.F. Filley, and G. Cumming. 1971. *Models of the Human Bronchial Tree*. J. Appl. Physiol. 31: 207-217.
3. Yeh, H.C., and G.M. Schum. 1980. *Models of Human Lung Airways and their Application to Inhaled Particle Deposition*. Bull. Math. Biol. 42: 461-480.
4. Phalen, R.F., M.J. Oldham, C.B. Beaucage, T.T. Crocker, and J.D. Mortensen. 1985. *Postnatal enlargement of human tracheobronchial airways and implications for particle deposition*. Anat. Rec. 212: 368-380.
5. Martonen, T.B., Y. Yang, D. Hwang, J.S. Fleming, 1995. *Computer Simulations of Human Lung Structures for Medical Applications*. Comput. Biol. Med. 25: 431-446.
6. Kitaoka, H., R. Takaki, and B. Suki. 1999. A Three-Dimensional Model of the Human Airway Tree. J. Appl. Physiol. 87: 2207-2217
7. Tawhai, M.H., A.J. Pullan, and P.J. Hunter. 2000. *Generation of an Anatomically Based 3D Model of the Conducting Airways*. Ann. Biomed. Eng. 28: 793-802.
8. Fleming, J.S., M.A. Nassim, A.H. Hashish, A.G. Bailey, J.H. Conway, S. Holgate, P. Halson, E. Moore, and T.B. Martonen. 1995. *Description of Pulmonary Deposition of Radiolabeled Aerosol by Airway Generation Using a Conceptual Three-dimensional Model of Lung Morphology*. J. Aerosol. Med. 8: 341-356.
9. Fleming, J.S., V. Sauret, J.H. Conway, and T.B. Martonen. 2004. *Validation of the Conceptual Anatomical Model of the Lung Airway*. J. Aerosol. Med. 17: 260-269.
10. Fleming, J.S., J.H. Conway, S.T. Holgate, A.G. Bailey, and T.B. Martonen. 2000. *Comparison of methods for deriving aerosol deposition by airway generation from three-dimensional radionuclide imaging*. J. Aerosol. Sci. 31: 1251-1259.
11. ICRP Task Group. 1994. *Publication 66. Human Respiratory Tract Model for Radiological Protection*. International Commission for Radiological Protection, Pergamon Press, Oxford.
12. Sauret, V., P.H. Halson, I.W. Brown, J.S. Fleming, and A. G. Bailey. 2002. *Study of the three-dimensional geometry of the central conducting airways in man using computed tomographic (CT) images*. J. Anat. 200: 123-134.

A Simple Electrical Equivalence Model of Intracranial Cerebrospinal Fluid Pulsatility: Design and Validation in Healthy Normals

Jieun Kim[a], Neil A Thacker[a]*, Paul A Bromiley[a], Stephen J. Payne[b] and Alan Jackson[a]

[a]Imaging Sciences and Biomedical Engineering, University of Manchester, Manchester M13 9PT
[b] Biomedical Engineering, University of Oxford, Oxford

Abstract. During the cardiac cycle a complex series of fluid shifts occur within the skull in order to protect the brain from the pressure variations which occur in the cerebral arteries. The extracerebral intracranial arteries dilate during systole generating a pressure wave within the cerebral spinal fluid (CSF). This pressure wave is dissipated by flow of CSF into the compliant spinal subarachnoid space and by direct transmission to the cerebral venous sinuses. This mechanism reduces the pulsatility of the pressure wave to which the brain is exposed during the cardiac cycle. The mechanism can be investigated using quantitative magnetic resonance phase imaging but results can be difficult to interpret due to the complexity of the interactions. We present a novel physiological model of this mechanism based on the concept of electrical equivalence. We tested the model in a group of 24 healthy normal volunteers. Analysis of individual subjects showed that the data contained in adequate information for reliable fitting. Groupwise analysis showed that the model described all of the statistically significant variation in the data. We conclude that this model forms a basis for the analysis of CSF flow studies.

1 Introduction

The advent of magnetic resonance imaging (MRI) which provides the ability to non-invasively and quantitatively image CSF and blood flow greatly increased the understanding of CSF dynamics and it has become clear that the constraints of the Monro-Kellie hypothesis result in a complex hydrodynamic mechanism which is required to compensate for the transient increases in cerebral blood volume that occurred during systole [1]. The complexity of this autoregulatory process, the variations in measurements and interpretation methods recommended by different groups [1, 2, 3] and the limitations on the number of measurements which can be obtained by MRI combine to make such phenomenological observations increasingly unsatisfactory. There is a clear need for a quantitative model of the mechanism to act as a substrate for principled analysis of this type of data. A full modelling approach would require a level of detailed knowledge relating to tissue structure and its properties that is unrealistic for practical application.

In this paper we construct a bulk parameter model based upon an analogy to an electrical circuit. We then evaluate the model using MRI data obtained from young healthy control subjects. The analogy used to model the biological system as an electrical circuit is shown in Fig 1. This paper represents the first evaluation of an electrical equivalent model using measured data from living subjects in comparison to previous work [4].

2 Materials and Methods

Electric Equivalent CSF Pulsatility Model

We assume that we can represent all time varying signals in the Fourier domain so that we can analyse the equivalent circuit at a fixed set of frequencies w. We can then analyse this circuit (Fig 1 (b)) using conventional means by specifying current flows and deriving equations describing current and voltage (I_i - flow (current), C_i - compliance (capacitance), R_i - Impedance (resistance), V_i - blood pressure(potential)). Equations which describe the model, Fig 1(b), are presented in Appendix and detailed analysis of circuit is presented in [5]. Here are equations we derived from the model:

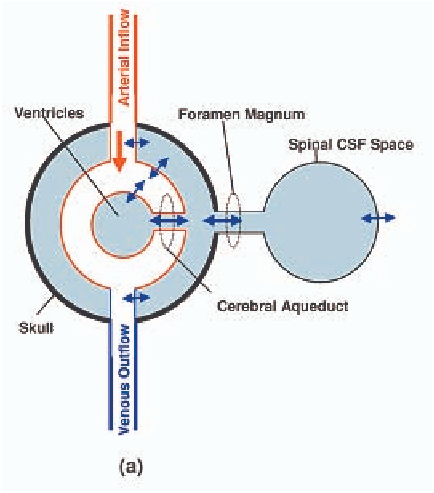
A statement of volume preservation, also known as Monro-Kellie Principle:

$$I_1 \; = \; I_2 \; + \; I_4 \tag{1}$$

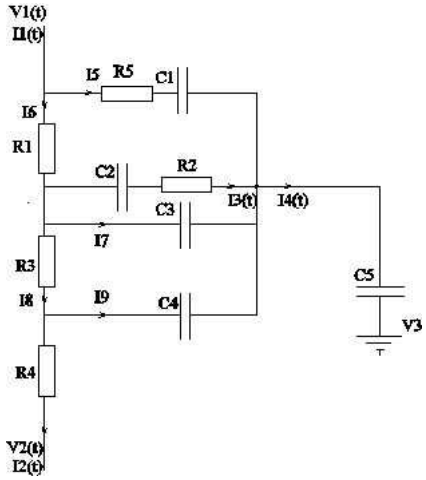
Constraint equation relating currents (flows):

$$(R_5 + D_{1w} + R_1)[I_3(D_{2w} \; + \; R_2)(D_{3w} \; + \; D_{4w} \; + \; R_3) \; - \; D_{3w}(I_4 \; - \; I_3 \;)D_{4w} \; - \; D_{3w}(I_4 \; - \; I_3 \; + \; I_2)R_3]$$

*Nail.Thacker@manchester.ac.uk



(a) Simplified anatomical model on which the electrical equivalence model is based.



(b) Electrical Circuits for the simplified biological model.

Figure 1. Biological system and electric current.

$$+ D_{3w}[I_3(R_2 + D_{2w}) + (I_2 + I_4)R_1](R_3 + D_{4w}) = 0 \tag{2}$$

where $\frac{1}{jwC_n} = D_{nw}$, all R's are real and all D's imaginary (all I's are complex);

We can build a suitable likelihood function from Eq. 2. Substituting Eq. 1 and re-organising terms, we obtain an equation of the form

$$\gamma I_3 - \alpha I_1 + \beta I_4 = 0 \tag{3}$$

where α, β and γ are complex variables found appropriately.

The above equations define the noise-free constraint. Applying the variational method to the constraint equation Eq. 3, the complex residual on the constraint for each Fourier amplitude w in the measured currents is;

$$F_w = \gamma I_{3w} - \alpha I_{1w} + \beta I_{4w}.$$

We then obtain the optimisation function;

$$-log(P) = \sum_w F_w^* F_w / var(F_w) \tag{4}$$

where the variance on the complex residual is given by

$$var(F_w) = \alpha^* \alpha \, \sigma_1^2 + \beta^* \beta \, \sigma_4^2 + \gamma^* \gamma \, \sigma_3^2$$

where σ_i are standard error for I_i.

Having determined the most likely parameters for the model we can now make corrections to the flow variables ΔI_n in order to enforce the constraint equations, in a way which minimises the change in the measurements consistent with the measurement errors:

$$\chi^2 = \sum_w \Delta I_1^* \Delta I_1 / \sigma_1^2 + \Delta I_3^* \Delta I_3 / \sigma_3^2 + \Delta I_4^* \Delta I_4 / \sigma_4^2.$$

Subjects and MR Image

MR image volumes were acquired from 24 healthy control subjects (18M, average 35yrs and 6F, average 24yrs). Velocity encoded PC-cine MR images were acquired for each region. Retrospective cardiac gating using vector electrocardiography (VECG) was used to generate 16 images spanning the cardiac cycle for each subject. All subjects

were scanned using a 3T Philips Medical Systems with a multi-element head coil. V_{enc} (velocity encoding) was set at 10cm/sec for the for AQ and FM, and 90cm/sec for the BA and CA. Scan time was was 15-30min including scout imaging, depending mainly on the subject's heart rate. Typical images at a single time point for each region are shown in Fig 2 (1-5).

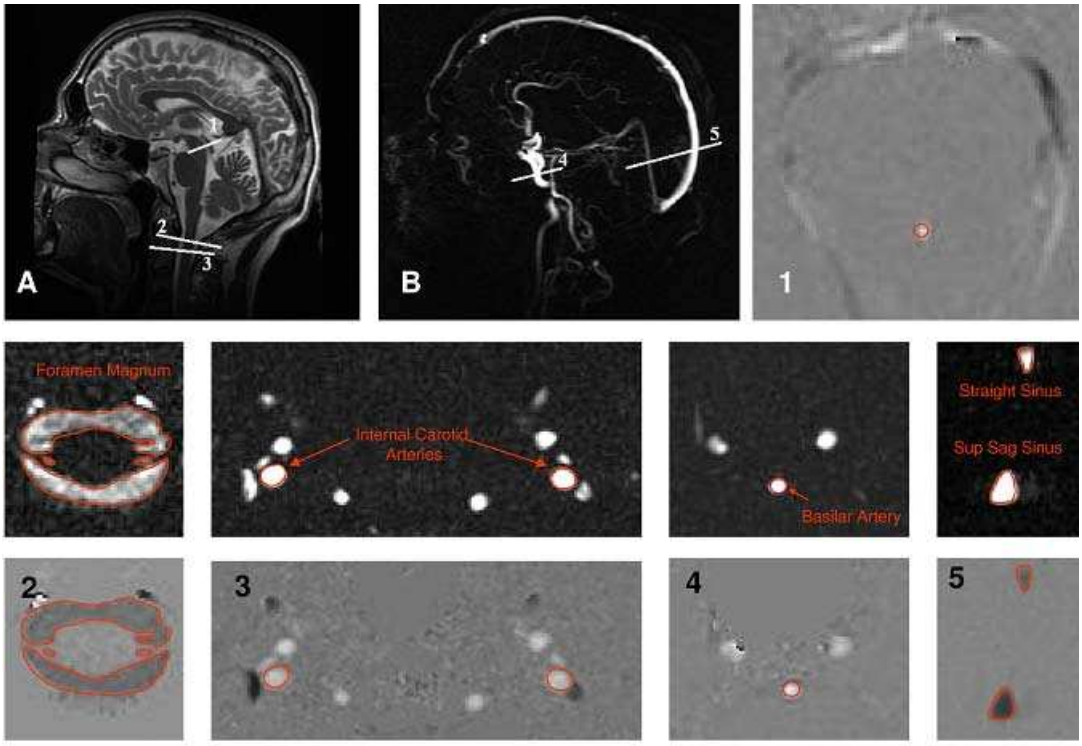


Figure 2. MR images. **(a)** Sagittal T1 scout image used to determine flow acquisition level and oblique axis for 1-AQ, 2-CA, 3-FM. **(b)** Phase contrast angiogram used to determine flow acquisition level for 4-BA, 5-SS. **1-5** Typical velocity encoded MR images of the vessels in which flows were measured. **1** AQ **2** FM **3** CA **4** BA **5** SS. The middle row shows corresponding modulus images

A quadratic function was fitted over the region of the cerebral aqueduct at each time point in the cardiac cycle. An approximate location for the centre of the aqueduct was provided manually and a 3x3 region was fitted for all subjects. The area under this curve was then integrated for all values of the function. The BA, CA and FM flows were estimated from the sum of flow values within the region of the arteries, defined by thresholding the modulus images to find the region of interest (ROI), as shown in Fig 2. Typical flow curves obtained from the MR images are shown as solid line in Fig 3.

3 Results

Given measured AQ flow (I_3), FM flow (I_4), and CAB flow (I1), Eq. 4 can be minimised as a function of the remaining parameters (R_2, R_3, R_5, C_1 - C_4). The estimated flows I_3 and I_4 can be corrected back to zero mean. The ratio of R1:R3 was fixed at 4:1 because the greater part of the pressure drop is known to occur in the arterial vessels [6]. We then normalised all remaining parameters to R_1. Typical input curves and error corrected model output flow curves for one subject are shown in Fig 3(a)-(c). Reproducibility study of 5 subjects suggested that the errors on the measured curves are 0.114 ml/sec for FM, 0.174 ml/sec for CAB and 0.00114 ml/sec for AQ. These errors were used to scale the Fourier terms in Eq. 2.

We took a group-wise approach, in which a sub-set of the data was selected, consisting of the 16 subjects who gave non-trivial solutions in terms of the C_3 and C_4 parameters for individual fits. The model parameters were then optimised simultaneously across this data group. This approach was based on the assumption that the physical parameters of the model do not vary greatly between normal individuals. Therefore, variations between individuals were implicitly assumed to represent differences in systemic behaviour. Thus, the group-wise approach provided additional information on the typical range of model parameters for normal subjects which further constrained the fitting process. The optimised parameters, normalised to R_1, are shown in Fig 4. The optimum parameters are given in Table 1.

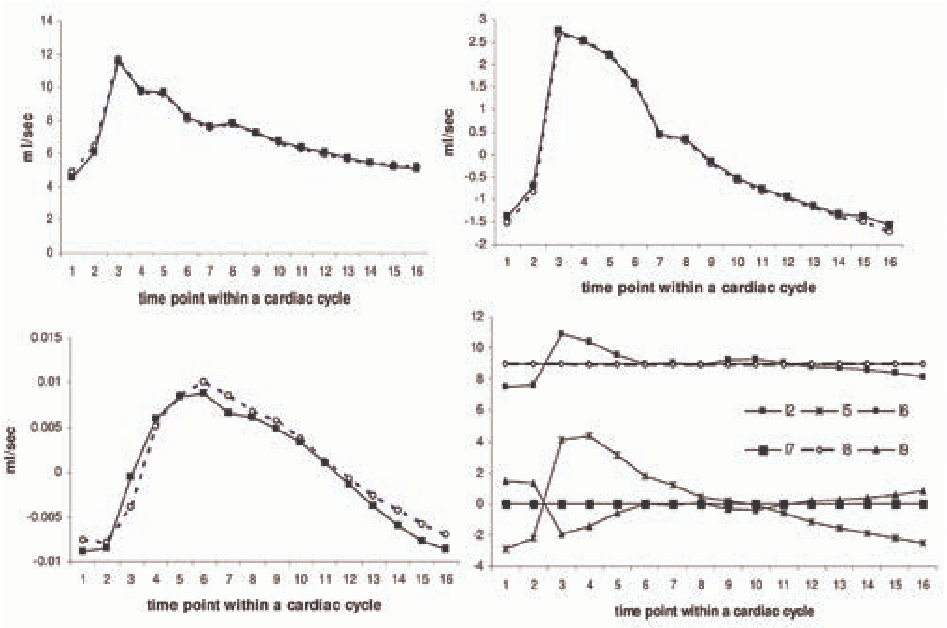


Figure 3. MR flow measurements (input to the model, dotted line) and estimated error-corrected flow output (sloid line) for one individual from the model at CAB (top left), FM (top right), AQ (bottom left), the remaining flow output from the model (bottom right).

Table 1. Fitted parameters from entire data set (arbitrary units).

Parameter	Physical interpretation	Value	Standard error
R_1	impedance of arterial capillaries	1.0	0.025
R_2	impedance of cerebral aqueduct	≈ 0.0	-
R_3	impedance of venous capillaries	≈ 0.0	-
R_5	impedance of arteries	0.01	8×1.8^{-3}
C_1	elastic capacitance of arteries	$\approx$large	-
C_2	elastic capacitance of ventricles	4.11	0.11
C_3	elastic capacitance of capillaries	≈ 0.0	5×10^{-3}
C_4	elastic capacitance of veins	271.017	18

4 Conclusions and Discussion

We have developed a simple electrical equivalence model which describes the interactions between intracranial arterial and CSF flow patterns observed in normal individuals. This is the first time that a model of this type has been tested with genuine data obtained from MR images of 24 normal subjects. The results from fitting the refined model to individual subjects indicated that, even with a markedly simplified anatomical model, the information content of the data is too low to provide stable estimates of the model parameters in the majority of cases. Therefore, a group-wise approach was adopted, in which the model parameters were optimised simultaneously over those data sets that returned non-trivial solutions for individual model fits. This procedure succeeded in producing a stable set of optimised parameters. Retrospective application of this optimised model to the data sets omitted from the group-wise fitting showed that, in all but one of the subjects, the model described all of the statistically significant variation in the data. This confirms that the difficulty encountered in fitting the model to individual data sets is related to the information content of the data, rather than an inability of the model to describe the whole range of normal variation. We therefore conclude that the optimised model provides a valid description of the pulsatile flow within the cranium in normal subjects. The failure of the model to fit in one individual case is most likely to represent anatomical or physiological variation within this individual which has not been specifically accounted for within the simplified model.

The simplified model allows quantitative derivations of seven parameters: 1) the arterial compliance, 2) brain compliance, 3) ventricular compliance, 4) venous compliance, 5) arterial impedance, 6) brain impedance and 7) impedance of the cerebral aqueduct. However, these seven parameters can only be measured up to a scale factor using Eq. 2 so we must select one and fix it to a nominal value. In order to estimate absolute values of the 7 estimated parameters it would be necessary to directly measure two pressures (arterial venous or CSF) within the system for each individual.

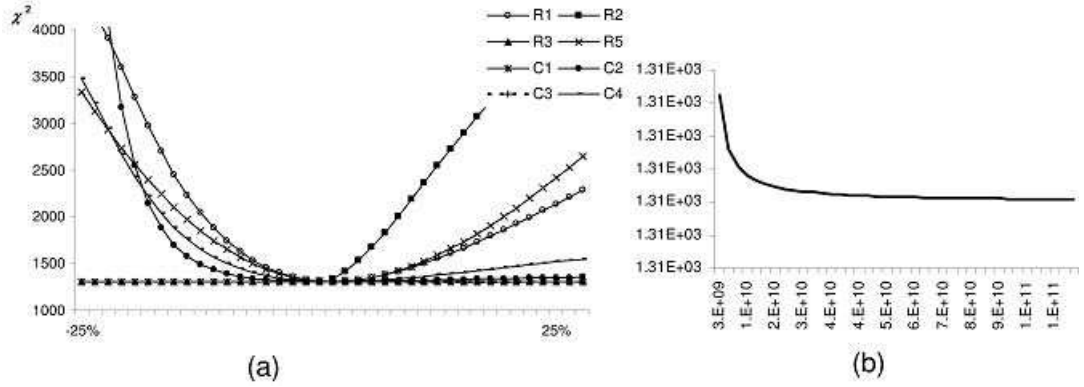


Figure 4. χ^2 values for fitted parameters for the entire group: The χ^2 has 762 degrees-of-freedom (16 measurements of 3 flow curves for 16 subjects, minus the 6 degrees-of-freedom of the model). **(a)** the abscissa is given as a percentage variation of the optimal parameter value. **(b)** shows the parameter fitting of C_1, where the abscissa is given as the actual parameter value.

Even under these circumstances it would be important to realise that the equivalence between compliance and capacitance is not straightforward. Compliance is the equivalence of permittivity, not capacitance, and to obtain compliance (the measure of tissue elasticity) would require scaling variables such as cross-sectional areas and volumes, which is impracticable at this time. Although the model is deliberately simplistic, it allows for potential extension to allow for more sophisticated description of anatomical and pathological variations. We conclude 1) that the model presented here provides a valid description of the interaction between into cerebral blood flow and CSF movements and 2) the model be used to support is to allow inclusion of additional anatomical and pathological variation which will be required to support its application in disease states.

References

1. D Greitz. Radiological assessment of hydrocephalus: new theories and implications for therapy. *Neurosurg Rev* 2004;27(3): 145-165, 2004.
2. G A Bateman. The reversibility of reduced cortical vein compliance in normal pressure hydrocephalus following shunt insertion. *Neuroradiology* 45:65-70, 2003.
3. T Miyati, M Mase, T Banno, T Kasuga, K Yamada, H Fujita, K Koshida, S Sanada and M Onoguchi. Frequency analyses of CSF flow on cine MRI in normal pressure hydrocephalus. *Eur Radiol*2003;13:1019-1024, 2003.
4. M Ursino and C A Lodi. Interaction among autoregulation, CO2 reactivity, and intracranial pressure: a mathematical model. *Am J Physiol* 274(5 Pt 2) H1715-28, 1998.
5. J Kim, NA Thacker, PA Bromiley, SJ Payne and A Jackson A Simple Electrical Equivalence Model of Intracranial Cerebrospinal Fluid Pulsatility: Design and Validation in Healthy Normals. *J Cerebral Blood Flow & Metabolism* (Submitted) 2006.
6. R M Berne and M N Levy. *Caridiovascular Physiology.* London: Mosby, 2001.

Appendix

Equations we obtain from Fig 1(b) by applying Kirchoff's laws:

Vertex Currents

$$I_1 = I_5 + I_6; \quad I_8 = I_9 + I_2; \quad I_6 = I_3 + I_7 + I_8; \quad I_4 = I_5 + I_3 + I_7 + I_9$$

Voltage loops

$$I_5 R_5 + I_5 \frac{1}{jwC_1} - I_3 R_2 - I_3 \frac{1}{jwC_2} - I_6 R_1 = 0$$

$$I_7 \frac{1}{jwC_3} - I_9 \frac{1}{jwC_4} - I_8 R_3 = 0; \quad I_3 \frac{1}{jwC_2} + I_3 R_2 - I_7 \frac{1}{jwC_3} = 0$$

Point to point voltages

$$V_2 - V_1 = I_6 R_1 + I_8 R_3 + I_2 R_4; \quad V_3 - V_1 = I_6 R_1 + I_3 \frac{1}{jwC_2} + I_3 R_2 + I_4 \frac{1}{jwC_5}$$

Independent Component Analysis Based Active Shape Model with Spatial Relations for Finding Correspondence

Zihua Su, Tryphon Lambrou, Andrew Todd-Pokropek

Department of Medical Physics & Bioengineering, University College London,
Malet Place Engineering Building, Gower Street, London WC1E 6BT, U.K.

Abstract. Statistical shape models use Principal Component Analysis (PCA) to describe the shape variations. However, PCA has the restriction that the input data must be drawn from a Gaussian distribution, and is only able to describe global decomposition. In recent years, Independent Component Analysis (ICA) has become a popular alterative for shape decomposition. Due to the local variations that ICA represents, the final optimal result usually turns out to be an invalid shape. In this paper, we will investigate the details of the ICA-ASM. With the consideration of the influence from neighbourhood points by using Markov Random Field (MRF), we overcome this drawback introduced by ICA. Our initial results show that our proposed method offers a better rate in obtaining a valid shape. From this, we can conclude that the MRF based ICA model provides improved results to the Bayesian based ICA model currently used.

1 Introduction

Geometric shape information plays a key role in many computer vision and image processing applications, especially in medical image analysis where many anatomical structures and organs can be identified and classified in terms of their unique shapes. The correspondence is such a critical step that usually comes before Procrustes alignment.

The advantage of Active Shape Model (ASM) is that it uses experiences from training data to judge the correspondence and shape outline. In recent years ICA has been introduced into ASM for its excellent performance on giving more accurate local variations and no restrictions on data set. Actually, ICA has become a more general data description method than PCA does. By using the Bayesian frame work [1], we can cast our problem into a Maximum a Posterior Probability (MAP) work. The final result shape is the optimal minimization. The problem is that using ICA the optimization method will face more local peaks than PCA does. Therefore, the result is more likely to be trapped in invalid shapes. We noted that some work has been done for eliminating this effect. By adding more artificial training datasets Wang et al [1] made their shape model more rigid and global. To some extent, they solve the problem of invalid shape, but that makes it harder to capture local variations. By using MRF theory [2], our method successfully conquers this drawback. We also note that some other approach has been made by R.H. Davies et al [3].

2 Method

2.1 ICA-Based Active Shape Model

ICA is a more general description of data format than PCA, since PCA can only represent the orthogonal condition. An example is shown in Figure 1. This may lead to the wrong description of datasets and results in shapes that are not plausible and sometimes can not generate the shape desired. Therefore, ICA as an alternative can potentially resolve these problems. PCA assumes the data to be Gaussian distributed and in most of cases that does not hold [4]. In addition, ICA gives a very convenient advantage, which PCA can not easily have, that the joint probability of all the components is equal to the product of every component's probability.

In our proposed method we use similar ranking algorithm to that presented in [5]. After that, we select the first t components to cover 95% of all variations.

Since we know that finding the maximum negentropy direction is equivalent to finding a representation in which mutual information is minimized and negentropy is the natural measure of non-Gaussianity. In that sense, the Gaussian model is not suitable for ICA components. Some other parametric approach like Laplace or double-exponential density, non parametric methods or mixture model can be adopted. For simplicity in manipulation, we choose the mixture Gaussian model.

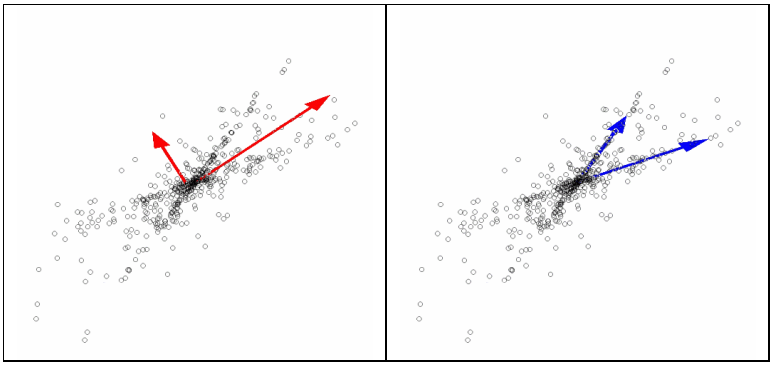


Figure 1. An example of PCA (left) and ICA (right), as we can see, ICA finds the real distribution format but PCA does not.

Here we propose a Gaussian mixture model for the ICA density expression. In this mixture model we are combining *M* parameterised densities and giving each one of them a weight. A frequently used algorithm for this optimization problem is the Expectation Maximization (EM) algorithm [4]. EM is an iterative method that finds the maximum by choosing a new guess to maximize the lower bound. Some of the optimization results from the real training sets are shown in Figures 2 and 3.

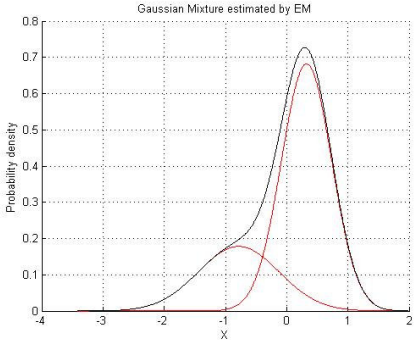


Figure 2. Gaussian Mixture estimated by EM.

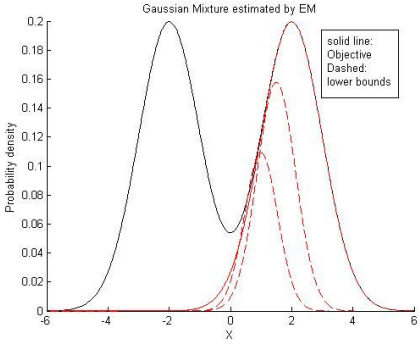


Figure 3. Intermediate steps of the EM algorithm.

2.2 Markov Random Field Regularization

Since ASM was developed, research has been carried out on its parameter optimization. Different methods have been used in order to compose objective functions. Among these methods, the Bayesian framework is widely employed due to its usage of prior knowledge. However, lack of consideration in relations between parameters makes the final optimal shape often invalid. Using ICA to substitute PCA makes this even worse, for achieving similar error levels ICA will need to use more components (i.e. eigenvectors) than PCA does. Wang, et al [1] has tried to add some rigid artificial variations to the eigenvectors in order to make the shape changes more global and reasonable. This, to some extent, solves the problem, but makes the model harder to capture local variations that often exist in practical cases and this will not work on all cases (Figure 4). A natural way to incorporating spatial correlations into Bayesian process is to use MRF as a priori model. Thus, we follow the four successive stages of the Bayesian paradigm: 1) Construction of a prior probability distribution *p*(d) for the deformation field *D* matching the template shape S_t from training data to source shape S_s. 2) Formulation of an observation model *p*(y|d) that describes the distribution of the observed shaped *Y* given any particular realization of the prior distribution. 3) Using Bayes theorem to combine the prior and observation model into the posterior distribution. 4) Drawing inference based on the posterior distribution.

At this point we provide some definitions from MRF theory in order to describe the probability distribution on a spatial arrangement of points. Neighbourhood system and cliques are the most important definition in MRF theory. Given a graph of *n* connected sites $S = \{s_i\}_{i=1}^{n}$, a neighbourhood system $N = \{N_s, s \in S\}$ is any collection of

subsets of S in which: i) $s \notin N_s$, and ii) $r \in N_s \Leftrightarrow s \in N_r$, then N_s are the neighbours of *s*. A clique *c* is a subset of sites *S* for which every pair of sites is neighbour. Let all cliques be denoted by *C*. For all $c \in C$ we assume that we have a family of potential function V_c.We may now define an energy function of any given configuration of *d* i.e. $U(d) = \sum_{c \in C} V_c$ .This leads to the definition of Gibbs measure. The Gibbs measure induced by energy function $U(d)$ is

$$p(d) = \frac{1}{z} e^{(-U(d)/T)} \quad (1)$$

where, *Z* is the partition function and *T* is a parameter referred to as temperature. The Gibbs measure maximizes entropy (uncertainty) among all distributions with the same expected energy. The temperature controls the "peaking" of the density function; normally we define it as 1 for simplicity. The normalizing constant may be impossible to obtain due to the problem of dimensionality but often we need only ratios of probabilities and the constant omitted. A theoretical result called Hammersley-Clifford gives the relation between MRF and Gibbs random fields and states that *D* is a Markov random field with respect to *N* if $p(d)$ is a Gibbs distribution with respect to *N* [7][8]. So, we need to specify potentials that induce the Gibbs measure in order to encompass MRF properties of *D* on the graph. More details are given in [2].

2.3 Prior Distributions

We construct energy function based on differences between neighbouring sites. We put this in a multivariate case then we have the general expression of energy governing the site-priors

$$U_{site}(d) = \sum_{i \sim j} \left\| d_i - d_j \right\|_p \quad (2)$$

where, $\left\| . \right\|_p$ is the p-norm, *p=2* in the 2 D case, and d_i represents the multivariate displacement of the *i*th site.

With *p=2* the energy function induces a Gaussian prior on the deformation fields. Neglecting regions with strong surface dynamics the local optimization becomes concave and the maximum likelihood estimate of the displacement at the *i*th site is taken as the mean of the neighbouring displacements. Given the statistical models (ICA-ASM) and the shape parameters $a = (a_1 a_2 ..... a_t)^T$, and pose parameters: scale s, rotation θ , translation T_x, T_y . The combining pose and shape parameter vector to be determined is represented by the following equation.

$$P = (s, \theta, T_x, T_y, a_1, a_2, \ldots, a_t)^T \quad (3)$$

2.4 Observation Models

The observation model *p*(y|d) describes the conditional distribution of the observed data *Y*. By specifying an observation model we may favour a mapping that makes correspondence between regions of similar boundary properties. We propose using only the edge information in the input image, which is denoted as *E* here. The edge image *E* is assumed to consist of one of the deformed templates, t_p corrupted by additive white zero mean Gaussian noise with standard deviation σ_n, i.e. $E = t_p$ +n. This leads to (similar as in [9]):

$$U_{ob} = \sum_{n=1}^{N} \sum_{m=1}^{M} E(x(P,n), y(P,n)) k_m \frac{1}{\sigma_{n,m}^2} \quad (4)$$

where, *x(P, n), y(P, n)* are the coordinates in the test image, *k* the template magnitude at any point which is assumed to be a constant and is chosen to be the maximum boundary response, and *N* is the number of marks on the boundary.

The posterior equation is given by

$$p(d \mid y) \propto \exp(-U_{total} / T) \tag{5}$$

where, $U_{total} = aU_{site} + (1-a)U_{ob}$, in which $a \in [0:1]$ weights the sensitivities of different parts. When applying simulated annealing the posterior is linked to the prior and the observation model by

$$p(d \mid y) \propto (p(y \mid d))\, p(d))^{1/T} \tag{6}$$

3. Results and Discussion

In our experiment, a collection of 28 slices of different brain MRI datasets was used, in which the corpus callosum were labelled manually. On each image, 36 landmarks were labelled. Since we did not have a large data set, a leave-one-out experiment is performed, by repeatedly training the shape model on 27 of the images and testing it on the remaining image. The start position is selected by differing from the mean shape, either on X and/or Y coordinates by 10 pixels. A comparison is made between ICA based Bayesian model and our proposed MRF based method. In Figure 4, it can be seen that the new shape model finds the boundary with correspondences more reasonably accurately and the result is not sensitive to the start position. For the PCA and ICA algorithms, 12 and 20 eigenvectors were used, respectively. In both cases 4 extra degrees of freedom (i.e. scale, rotation, and two translations) were used.

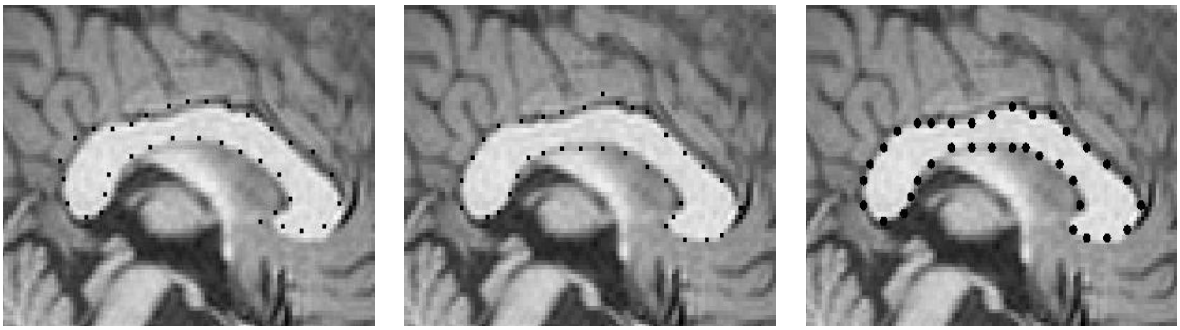


Figure 4. Three of the results: Bayesian ICA (left), Bayesian PCA (middle), MRF (right). We can see that MRF achieves better performance.

In order to evaluate the two methods accurately, we calculate the overall displacement of searched shape to the manually labelled shape (Gold Standard) for each test image. The distance of two shapes is defined as the sum of all absolute distances between corresponding points. We calculate MRFASM (our proposed method) and BAYEASM (original ASM). Then we calculate the improvement (*m*)

$$m = (BAYEASM - MRFASM)/BAYEASM \tag{7}$$

This is shown in Figure 5 (a) and (b). In both the graphs, the x- coordinate is the index of the test images, and the y-coordinate is its corresponding percentage improvement value. Figure 5(a), presents the comparison in terms of improvement (*m*) between the standard PCA Bayesian and our proposed MRFASM model. Figure 5(b), presents the comparison between ICA Bayesian and our MRFASM model. We can see from the Figure 5 (a) and (b) that our

method offers improvement between 8-24% for all the images on the testing dataset. Comparing the equations provided in [1] and equation (5) of our method, we can find that MRF introduces a stronger prior distribution by considering spatial relation between neighbour points. The use of this term is actually smoothing the cost function that can be the reason of the advantages.

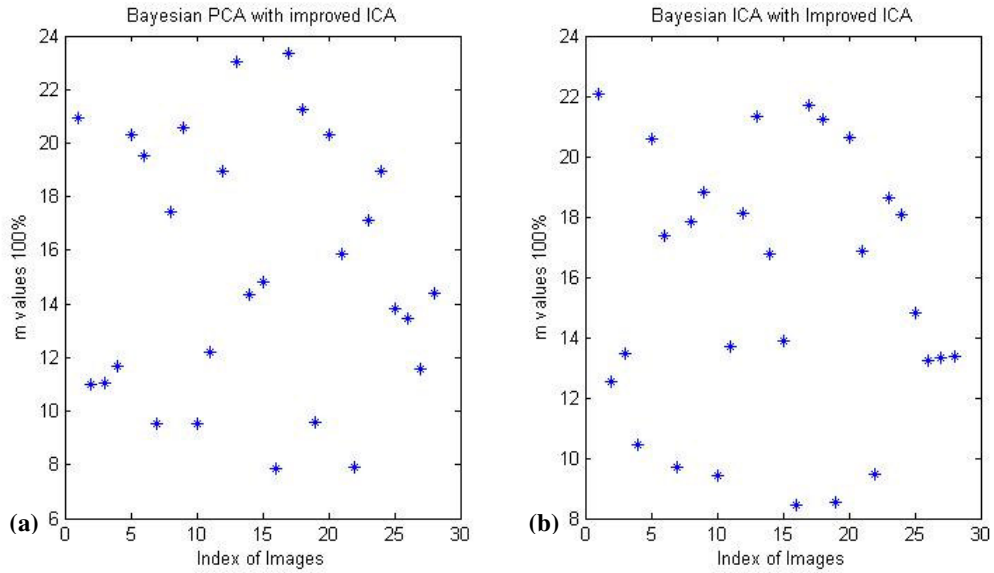


Figure 5. Shape to shape percentage distance difference, X direction is index of images; Y direction is percentage of improvement of ASM.

4. Conclusion

By using MRF, we provide more restrictive relations between parameters to make our ICA-based model converge more easily into a valid shape. Our initial results show that our proposed method offers a better rate in obtaining a valid and accurate shape. From this, we can conclude that the MRF based ICA model provides improved results to the Bayesian based ICA model currently used. Further work will include a larger dataset for the 2D case, as well as extending the technique into 3D, and different organs.

References

1. Y. Wang, L.H. Staib, "Boundary finding with correspondence using statistical shape models", Proceedings of IEEE, Conference on Computer Vision and Pattern Recognition, 338-345, 1998.
2. S.Z. Li, "Markov random field modelling in image analysis", Computer Science Workbench Series, 323, 2001.
3. R.H. Davies, C.J. Twining, T.F. Cootes, J.C. Waterton, C.J. Taylor, "A minimum Description Length Approach to Statistical Shape Modeling", IEEE Transactions on Medical Imaging, 21(5):525-537, 2002.
4. M. Uzumcu, A.F. Frangi, M. Sonka, J.H.C. Reiber, B.P.F. Lelieveldt, "ICA vs PCA Active Appearance Models: Application to cardiac MR Segmentation", Medical Image Computing And Computer-Assisted Intervention - MICCAI 2003, PT 1 2878: 451-458, 2003.
5. M. Uzumcu, A.F. Frangi, J.H.C. Reiber, B.P.F. Lelieveldt, "Independent component analysis in statistical shape models", Processing of the SPIE, Medical Imaging 2003: Image Processing, PTS 1-3 : 375-383 2003.
6. A. Dempster, N, Laird, D.Rubin, "Maximum likelihood for incomplete data via the EM algorithm", Journal of the Royal Statistical Society, 39:1-38, 1977.
7. J.M.Hammersley and P.Clifford, Markov field on finite graphs and lattices, 1971.
8. D. German, "Random fields and inverse problems in imaging", Lecture Notes in Mathematics, Springer-Verlag, 113-193, 1990.
9. L.H. Staib, J.S. Duncan, "Boundary finding with parametrically deformable models", IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(11):1061-1075, 1992.

Alignment of Maxillofacial CT Scans to Stone-Cast Models Using 3D Symmetry for Backscattering Artifact Reduction

Murat Balci, Mais Alnasser and Hassan Foroosh*

Computational Imaging Laboratory, University of Central Florida

Abstract. In this paper, we present an approach to accurately align the CT scans of a patient to a stone-cast model of his/her mandible or maxilla, and use the result of registration to clean up the patient's scans from artifacts and defects. The proposed approach assumes that the maxillofacial features are roughly symmetric with respect to a 3D plane. Then 3D volumetric models of both the patient and the stone-cast are reconstructed from the input data using a marching cube algorithm. The planes of symmetry are extracted using an improved Extended Gaussian Images method. After an initial alignment of the two volumes guided by the plane of symmetry due to 3D homology, we minimize a global cost function that depends on the sum of square differences(SSD) of patient data with the stone-cast model to finally recover the rigid transformation between the two scans.

1 Introduction

In dentistry, implant dentistry in specific, there is a need for 3-D visualization of internal anatomy. The internal visualization is mainly based on CT scanning technologies. As discussed in [1] the most important technological advancement which dramatically enhanced the clinician's ability to diagnose, treat, and plan dental implants has been the CT scan. Advanced 3-D modeling and visualization techniques permit highly refined, accurate assessment of the CT scan data in a manner, which exceeds information gleaned from film alone [1–3]. Since the advent of the Conebeam CT technology, which is a low dose oral and maxillofacial scanner, the number of dental CT scanners has been growing exponentially. The strongest barrier in taking a CT scan is the dose of x-ray exposure, which is not always justifiable in oral surgery cases. Conebeam scanners have an average of 10 times less radiation than regular medical CT scanners [4]. However, this comes at a price that the resulting scans are not free of defects.

In addition to imperfections of the instrument and the imaging process, it is not uncommon to encounter other unwanted artifacts in the form of bright regions, flares and erroneous pixels due to for instance dental bridges or metal braces. Currently, removing and cleaning up the data from acquisition imperfections and unwanted artifacts is performed manually, which is as good as the experience level of the technician. On the other hand the process is error prone, since the editing process needs to be performed image by image.

We attempt to address some of these issues by using stone-cast models of patient's dental imprint. Stone-cast models were originally used by dentists to make complete or partial dentures. It is suggested herein that the CT scan of such stone-cast model can be used to automatically guide the cleaning of patient's CT scans from defects or unwanted artifacts. The key, however, to such application would be an accurate registration of the two CT scan data, i.e. the patient's maxillofacial CT scans and that of the stone-cast model.

The problem is rather challenging since, in some cases it is a priori difficult to exactly identify which parts in the patient's scans are soft tissues, nerves, or bones. Prior radiodensity scales defined by Hounsfield units often give only approximate solutions, and are not always available in the patient's DICOM file. In order to solve this problem, we propose that the registration process can be drastically simplified and improved, if we use the fact that the maxillofacial features are roughly symmetric. Symmetry is a very strong cue that has been used in the past in imaging problems for somewhat opposite application of recovering 3-D models [5, 6]. Once the two scans are accurately registered the cleaning process can be performed by simple masking and morphological erosion.

Next section describes the proposed method. Section 2.1 explains the method proposed for identifying the plane of symmetry in the CT scans. Section 2.2 outlines how the estimated planes of symmetry can be used to guide the registration between the patient's scans and that of the stone-cast model. Section 2.3 is describing the refinement by optimization step and Section 2.4 proposes an algorithm to detect outliers of the maxillofacial CT scan data. In section 3, we illustrate the application of our algorithm to real CT scan data.

*Email : balci,nasserm,foroosh@cs.ucf.edu ,Research is supported by UCF I2Lab

2 Proposed Framework

The proposed method includes the following steps:

- Estimation of the plane of symmetry for both patient's and stone-cast CT-scans.
- Alignment of the planes of symmetry
- Alignment of scans by searching among transformations that would fix the plane of symmetry.
- Refinement using global minimization based on sum of square differences(SSD).

2.1 Extraction of Planes of Symmetry

A natural solution to estimate the planes of symmetry for the patients scan data and that of the stone-cast model might appear to be based on minimizing an error defined by 3D Homography. However, we found that such approach would be computationally expensive and would require a manual initialization. Therefore in this section we propose an approach based on working directly on the volumetric data. Similar to [7] our method makes use of extended Gaussian images (EGI) [8] to identify the 3D reflection plane. EGI have been proven to be very useful tools in such applications as pose estimation, orientation, and recognition [9–14]. The basic idea is exploiting the observation that in practice, if an object is symmetric, then so is its EGI. Therefore, by sampling the EGI and representing its histogram one can directly identify a point on the plane of symmetry as well as the normal to the plane. Sun and Sherrah [7] proposed a discrete approach to this problem. The first step in their algorithm is to tessellate a sphere in order to use it as an orientation histogram. A search is then performed near the principal axis directions for the strongest symmetries by performing correlation operations on the histogram. A unique labeling for the triangular facets (bins) resulting from the tessellation is required and thus also an efficient algorithm to determine the facet in which a given normal lies.

We avoid the tessellation, the labeling, and the need for an efficient search algorithm by simply parameterizing the unit sphere uniformly. We then create a two-dimensional orientation histogram by means of a simple table, where the indices i and j are determined as following:

$$i = \text{ floor}(u_1 n_y)\ ,\ j = \text{ floor}(u_2 n_x) \tag{1}$$

such that $0 \leq u_1, u_2 \leq 1$, n_x is the x-resolution and n_y is the y-resolution. Since we are dealing with symmetric objects a resolution of less than square root of the number of normals is often sufficient. Assuming that the CT-scans are triangulated, with all normals computed and depicted by two angles for each triangle in the facial model, the normal $\mathbf{n} = [\mathbf{n_x}, \mathbf{n_y}, \mathbf{n_z}]$ is computed, and hence its direction, depicted by the two angles φ and θ, would be given by:

$$\theta = \arccos(n_z)\ ,\ \varphi = \arctan\left(\frac{n_y}{n_x}\right) \tag{2}$$

and are then determined as following:

$$u_1 = \frac{1 - n_z}{2}\ ,\ u_2 = \frac{\varphi}{2\pi} \tag{3}$$

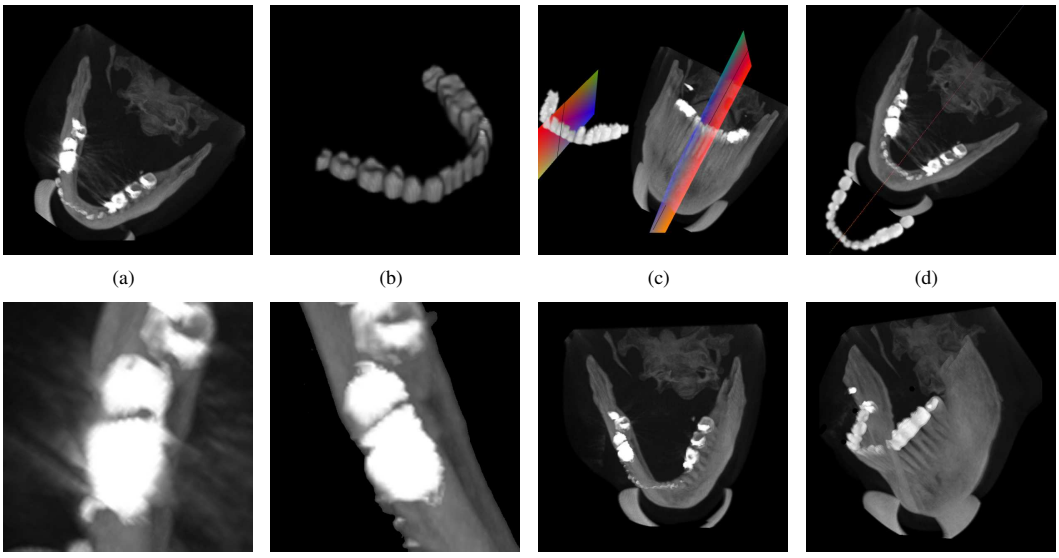
Once the orientation histogram is populated, a search is conducted for the largest count of normals. The corresponding indices i and j are retrieved and the xyz-coordinates of the most common normal **n** is determined using the following inverse relations:

$$\begin{aligned} u_1 &= \frac{i}{n_y}\ ,\ u_2 = \frac{j}{n_x} \\ \varphi &= 2\pi u_2 \\ n_z &= 1 - 2u_1 \\ r &= \sqrt{\max(0, n_z^2)} \\ n_x &= r\cos\varphi, n_y = r\sin\varphi \end{aligned} \tag{4}$$

Where r is the radius. The equation of the symmetry plane can now be found using the normal **n**, and the centroid **c** of the model, as the set of points that satisfy:

$$\mathbf{n}^T(\mathbf{p} - \mathbf{c}) = 0 \tag{5}$$

where **p** is a point on the symmetry plane. We found excellent results by applying this approach.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 1. (a) Original Patient Model (b) Corresponding Stone Cast Model (c) Initial Models with planes of symmetry (d) Models after planes of symmetry are matched (Top View) (e) A section of zoomed Patient Data (f) Partial zoom view of corrected data by use of stone-cast matching (g) Patient Model after clean up (h) Stone Cast model and cleaned up patient data merged together.

2.2 Matching Symmetry planes and 2D Search

Once the planes of symmetry are found in both the patient scans and the stone-cast scans, the two volumetric models can be pre-aligned. This is performed automatically by overlaying the two planes. In practice, this would provide an extremely good initialization, up to an unknown transformation with three degrees of freedom, i.e two translation parameters and one rotation.

After planes of symmetry are matched, we initially assume that the rotational misalignments are negligible. This reduces the search space for registration to pure 2-D translation. The correct translation parameters are then obtained by searching for 2D translation parameters that would minimize

$$\mathbf{T}^* = \arg\min \sum_{T} \|\boldsymbol{D} - \mathbf{T}\boldsymbol{S}\|^2 \tag{6}$$

where **D** is the is the original patient CT scan data set, **S** is the related stone cast CT scan data set, **T** is a translation matrix, and *T* is the space of all possible translations within the planes of symmetry.

2.3 Local Optimization

The previous step would essentially provide an almost perfect solution to the problem. However, for medical applications, accuracy of alignment plays a very important role. Therefore after the initial alignment described above, we perform a local optimization in 3D, within some neighborhood to refine and adjust for minor rotational and translational errors. This is done by performing the following optimization step

$$\mathbf{P}^* = \arg\min \sum_{P} \|\boldsymbol{D} - \mathbf{P}\mathbf{T}^*\boldsymbol{S}\|^2 \tag{7}$$

where $\mathbf{T}^*$ is calculated at the 2D search step, **D** is the original patient CT scan data set, **S** is the related stone cast CT scan data, *P* is the space of all possible transformations, **P** is a rigid transformation from the space of all transformations *P* in some neighborhood, and $\mathbf{P}^*\mathbf{T}^*$ is the registration result.

2.4 Identifying the Outliers of Maxillofacial CT Scan Data

After the registration step is completed, one can use the registration result in many ways; to prune the patient CT scan data, to overwrite the patient CT scan by corresponding stone-cast CT scans,etc. This usage of registration solves the

problem of outliers which will be segmented out by the stone-cast registration. However for the outliers that are within the stone-cast region or in the absence of stone-cast models, there is a need for outlier detection. We can detect outliers, and we can use the stone-cast values to fill the gap, which will be caused by removal of outliers.

We propose an automatic thresholding algorithm which extends the isodata thresholding algorithm [15,16]. Originally, isodata algorithm works on the principle of evaluating a unique threshold τ for a given histogram assuming

$$\tau = \frac{\mu_0 + \mu_1}{2} \tag{8}$$

such that μ_0 and μ_1 are the means of components defined by the threshold. Since they are not known before having the threshold, an iterative algorithm was proposed:

- Select an initial threshold
- Calculate two means of the segments defined by current threshold
- Set new threshold to be average of calculated means
- Repeat until the threshold value converges

which corresponds to finding the threshold τ that minimizes the following error function e^2:

$$e^2 = \int_0^{\tau} (i - A)^2 h(i) di + \int_{\tau}^{N} (B - i)^2 h(i) di \tag{9}$$

where i indicates the gray level of the pixels (from 0 to N), h(i) is the histogram weighting for every ray level, A and B are the binary values of each of the two components, and τ is the threshold. For proof, please refer to [16].

The proposed thresholding algorithm is based on the idea that a good threshold should still be a good candidate for different neighborhoods defined by a window mask. So we propose to extended isodata thresholding algorithm by applying a short length windowed version through the histogram of complete patient CT scan data. Each resulting short-length threshold candidate increases the corresponding gray level's vote count. And this approach is repeated with increasing window sizes as well to provide multi-scale solution. At the end, the threshold with maximum votes is announced as the best current threshold for the given histogram. With the assumption that our data is composed of space (background), non-bone (tissue,etc), bones and outliers which are generated due to imperfections or dental add-ons, the problem converts into finding 3 good thresholds by repeating the above process. Moreover due to nature of the data, first threshold which corresponds to the space is already known. After a threshold is found, the histogram is cropped accordingly and the next threshold is searched on the rest of the histogram. This approach lead us to finding multiple good thresholds.

3 Experimental Results

We have evaluated our results experimentally on several cases of corresponding patient/stone-cast CT scans. We show one set of results below. Figure 1(a) shows a view of the 3D patient model. And Figure 1(b) shows a view of the corresponding stone cast model used. While Figure 1(c) shows the initial situation after planes of symmetry are extracted, Figure 1(d) show the system after the transformation to match the symmetry planes between the stone-cast model and the patient's CT scan. Once the two scans are registered the patient's data can be improved by removing the unwanted artifacts or defects. Figure 1(e) shows defects of the input data in detail, and Figure 1(f) shows the same region cleaned up by use of the algorithm. The corresponding 3D view after clean up is shown in Figure 1(g). Merging these two models also gives us a better result for modeling purposes. Figure 1(h) shows two models merged after registration and clean up processes.

4 Conclusion

We present an efficient algorithm to register two CT scans by using the symmetry of the scanned areas. The approach is proven to be very useful for registering and removing defects of maxillofacial scans of patients that require oral surgery, but can not be subjected to high dose radiation, in which case Conebeam scanners would play a very important role. The method is already being evaluated at low scale clinically by physicians, and is proven to produce extremely satisfactory results by the end users.

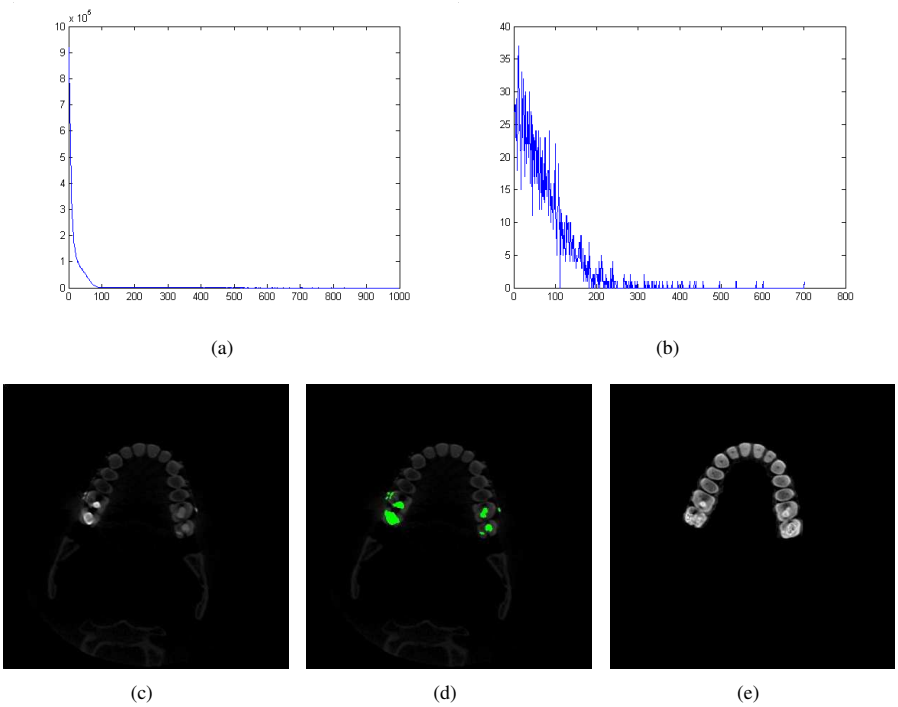


Figure 2. (a) Histogram of complete CT Scan data set (b)Part of the histogram cut by outlier threshold. Please note that there are just a few hundreds of voxels are marked as outlier in $\sim$ 70 millions (c) An example of sample data slice. (d) Corresponding outliers marked as green. (e) Result after outlier removal is shown. Note that the color range is also improved by outlier removal as well.

References

1. S. G. DMD. "Ct scan technology. an evolving tool for avoiding complications and achieving predictable implant placement and restoration, international magazine of oral implantology 1." pp. 6–13, 2001.
2. "Diagnostic possibilities of cone-beam computed tomography in the facial skeleton, international congress series 1268." pp. 1179–1183, 2004.
3. A. Gahleitner & B. Podesser. "Dental ct and orthodontic implants: imaging technique and assessment of available bone volume in the hard palate. european journal of radiology 51." pp. 257–262, 2004.
4. E. Asante, C. Eberhart & J. I. et al. "Staff utilization survey. sudbury, ma: American healthcare radiology administrators." pp. 15–27, 2001.
5. Z. Zhang & H. Tsui. "3d reconstruction from a single view of an object and its image in a plane, in proc. of fourteenth international conference on pattern recognition." **2**, pp. 1174–1176, 1998.
6. A. Francois, G. Medioni & R. Waupotitsch. "Reconstructing mirror symmetric scenes from a single view using 2-view stereo geometry, in proc. of 16th international conference on pattern recognition, pp. 12-16." 2002.
7. C. Sun & J. Sherrah. "3d symmetry detection using the extended gaussian image, ieee transactions on pattern analysis and machine intelligence, vol. 19,no 2." 1997.
8. B. K. P. Horn. "Extended gaussian images." *Proceedings of the IEEE* **17**, 1964.
9. M. Kampel & R. Sablatnig. "Range image registration of rotationally symmetric objects, in brandle n., (ed.), proc. of the computer vision winter workshop, rastenfeld, austria." pp. 69–77, 1999.
10. A. Goshtasby, G. Stockman & K. Rohr. "2-d and 3-d image registration: A tutorial computer vision and pattern recognition." 2004.
11. A. Goshtasby. *2-D and 3-D Image Registration For Medical,Remote Sensing, and Industrial Applications*. Wiley Interscience Publications, 2005.
12. L. G. Roberts. "Machine perception of 3-d solids, ph.d. thesis, mit." 1963.
13. M. A. V. Petra A. van den Elsen, Evert-Jan D. Pol. "Medical image matching -a review with classification." 1993.
14. R. A. Robb. "Biomedical imaging, visualization and analysis, wiley-liss press." 2000.
15. T. Ridler & S. Calvard. "Picture thresholding using an iterative selection method in ieee trans. system, man and cybernetics, smc-8." 1978.
16. A. Magid, S. Rotman & A. Weiss. "Comment on 'picture thresholding using an iterative selection method' in ieee trans. system, man and cybernetics, vol. 20, no. 5." 1990.

Improved Statistical Edge Detection Through Neural Networks

Ian Williams[a]*, David Svoboda[b], Nicholas Bowring[a], and Elizabeth Guest[c]

[a]Department of Engineering and Technology, Manchester Metropolitan University, UK
[b]Centre for Biomedical Image Analysis, Faculty of Informatics, Masaryk University, CZ
[c]School of Computing, Leeds Metropolitan University, UK

Abstract. A novel multi-statistic edge detector is presented. This edge detector utilizes a number of different statistical tests to analyze different properties across mask based regions within an image. The results from these various tests are then combined using a linear feed-forward Artificial Neural Network to create an optimum grey-scale edge detected image. The performance of this novel technique is objectively assessed against the traditional Canny filter using an adapted figure of merit calculation which works successfully on grey-scale images (GFOM). This evaluation illustrates how the novel filter can outperform the more traditional edge detectors on images where texture is a more significant edge discriminator than intensity, namely in histological image data.

1 Introduction

Edge detection is a low level process in many computer vision and image understanding applications. Edges are vital to region segmentation, object recognition and robot navigation processes and their success is directly dependent on the accuracy of the edge detection. This has led to the development of many algorithms for edge detection with the goal of producing the optimum results for a range of applications and images. Image edges are usually represented as discrete changes in the intensity profile of neighbouring pixels. Where this is assumed the derivative of the grey levels can be computed to determine the location of an edge. Many early detectors were derivative based filters, for example Roberts, Sobel, and Prewitt [1]. These perform effectively on synthetic images where the intensity change is clear, or images with very little noise. If the edge intensity difference is corrupted with noise or eliminated completely by image texture, the derivative based edge detectors perform less well. In such situations Gaussian smoothing techniques can be applied to reduce the effects of noise (Canny [2]). Gaussian smoothing however removes information from the image which may be required by a later process. Furthermore in the case of a large smoothing kernel it can also introduce problems relating to the true location of the edge. This effect can be reduced by the use of non-linear diffusion filtering [3] as a pre cursor to any gradient based edge detection technique but this does introduce a measure of user subjectivity to the results. To avoid these problems we can look for edges differently. An edge within a noisy or cluttered image can be seen as an interface between two regions of differing texture as opposed to differing intensities. This allows the edge to be located without using either derivatives, or image smoothing techniques.

2 Previous Work

Although not the most common approach used in edge detection, texture based edge filtering is becoming more prominent, particularly when applied to images corrupted with noise. DeSouza [4] illustrated a comprehensive analysis of five statistical tests for the detection of one dimensional edges, particularly rib structures in X-Ray imagery. The theoretical use of many statistical tests within 2D images has since been described by Bovik et al [5]. This work, although largely theoretical and so lacking in a full analysis of the results, illustrated the problem of high computational requirements for ranking statistical tests. Similar tests were described by Beauchemin et al [6] to overcome low image SNR, and also by Huang [7] to overcome the blurring effect evident with gaussian smoothing filters. Two dimensional edge detection using image region statistics were first applied by Guest [8], who introduced a novel pixel mask edge detection filter to accurately locate edges in noisy histological Mouse Atlas MA images. Furthermore the author identified the limitations with existing edge detectors on texture based edges. Fesharaki and Hellestrand [9] used a similar technique to successfully detect edges in both noise-free and noise-corrupted images. Comparisons to gradient techniques have since illustrated how statistical methods have a robust performance in the presence of noise, notably Kundu et al [10] and Hou [11], who both illustrate how statistics can outperform derivative methods on images corrupted noise. Lim [12] [13] has illustrated how a modified T-test performs well on images with added noise, this however could be outperformed by the Kolmogorov Smirnov test in intense noise images. Work by Bowring et al [14] has further indicated the possibility of producing images superior to Canny using a novel region based statistical technique. Recent work by Williams et al [15] has illustrated how this approach can be extended to multiple image regions of varying scales and artificial neural networks (NN) to further improve the edge detection quality.

*e-mail: i.williams@mmu.ac.uk

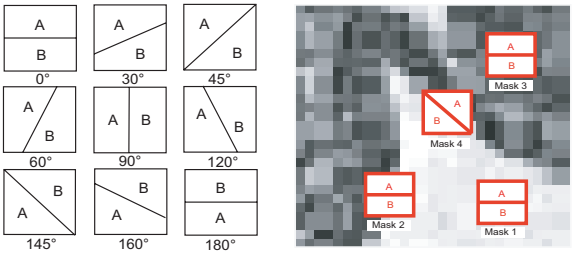


Figure 1. (a) An example of nine statistical masks, and (b) four different masks superimposed to an image. Each mask is divided into two regions *A* and *B* surrounding the central pixel. The orientation of the central line in the mask indicates the possible edge direction and also the mask angle. Masks 1 and 3 indicate a position where there is no edge. Mask 2 is located on the edge correctly although at a wrong direction. Mask 4 fits the edge precisely.

3 Method

The statistical filter used here was initially described by Guest [8] and further developed by Bowring et al [14]. The principle of this novel filter is briefly illustrated in Fig. 1, which shows a section of a histological *Mouse Atlas Image* (MA) [8] with a clearly defined texture edge. The main idea is based on the two-sample statistical test evaluation over two image regions *(A and B)* and a consecutive decision whether the inspected area contains an edge or not. For a more detailed description of the workings of this novel statistical method please refer to the work of Bowring et al [14]. Before we describe the practical implementation of the multi-test filter we will first briefly illustrate the workings of the particular statistical tests. In general, there are many useful statistical tests, however here we focus only on those which gave the best results in practice. Each is briefly classify as belonging to one of three main types: Mean based, Variance based, and Ranking Distribution based.

Mean Based Tests The *Student's T test* is a common mean based statistical test, based on the hypothesis that two distributions will have the same or a similar mean value. The two-sample T test for two equally sized regions is given as:

$$T = |\bar{x}_A - \bar{x}_B| \sqrt{\frac{N-1}{s_A + s_B}} \tag{1}$$

where $\bar{x}_A$ and $\bar{x}_B$ is the mean of region *A* and *B*, respectively. N is the number of pixels in one region with s_A and s_B being the variances of the two regions *A* and *B*.

The second mean based test used in this work is the *Difference of Boxes* test. This calculates the difference in the means of the two regions of the statistical mask. Usually two varying size mean filters are used [16], however here we simply use the mean of the two different mask regions *A* and *B*, given as $D = |\bar{x}_A - \bar{x}_B|$.

Variance Based Tests In some circumstances, particulary on biomedical images, the mean of the two mask regions does not differ significantly and a more robust test for statistical edge detection is to analyze the variance of the regions. The *Likelihood Ratio test* based on the work of DeSouza [4] is defined as:

$$L = -N \cdot \log_e \frac{4 v_A^2 v_B^2}{v_{A\cup B}^4}, \quad \text{where} \quad v_A^2 = \sum_{x \in A} x^2 - \frac{1}{|A|} \left(\sum_{x \in A} x \right)^2 \tag{2}$$

Here v_B and $v_{A\cup B}$ are defined analogously.

Ranking Distribution Tests The last statistical test used here checks neither the variance or the mean of the two regions, but simply via sorting or ranking checks for any difference in the two distributions. The test presented here is the *Mann Whitney U test*. It checks the hypothesis that the two data sets under evaluation are taken from the same distribution. The statistical value $U = \min(R_A, R_B)$ corresponds to a rank score which is calculated for both data sets. Here

$$R_A = \sum_{x \in A} \left(\sum_{y \in B; y < x} 1 \right) \quad \text{and} \quad R_B = \sum_{x \in B} \left(\sum_{y \in A; y < x} 1 \right) \tag{3}$$

refer to the data originating from mask *region A* and *region B*, respectively. Measure U is the overall statistical significance relating to the minimum value between R_A and R_B.

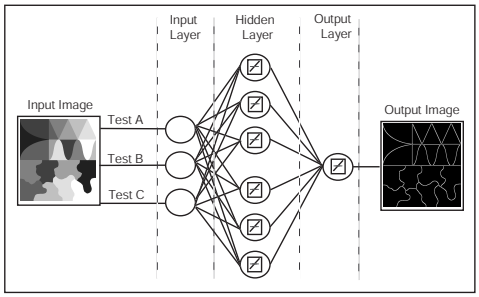


Figure 2. The workings of the multi-test filter. The inputs for the neural network are extracted from three different statistical tests applied to the same image (Tests A,B,C) . The linear output of the network is used to reconstruct the output image pixel by pixel.

3.1 The Multi-Statistical Test Filter

The statistical filter described Bowring et al [14] allows for the implementation of a single test to be computed at any time. Since we are assessing the textural properties of different image regions we have observed that if the regions have greatly differing mean values then the Difference of Boxes test or the Student's T Test will perform the best. If however we have regions with similar means but differing variances, both of these tests will fail and the Likelihood Ratio test would be much more appropriate. A performance analysis of these single tests for use on histological images was described by Svoboda et al [17] who indicated that no single statistical test alone could perform well across all types of texture edges, so illustrating the need for a multi-test approach. This has lead to the development of the multi-test filter which allows the combning of several statistical tests, in parallel, through the use of a neural network.

In this paper the neural network accepts several statistical edge detectors as its inputs. In practice only three tests of differing properties (i.e. variance based, mean based, and rank based) were required (see Fig. 2). The hidden layer of the network has a set of nodes each of which is identical in purpose and sum the inputs from each previous layer node. The final layer of the network consists simply of one linear neuron. This architecture allows us to have a single linear output from the network that once normalized becomes the appropriate output pixel value for the edge image. The training data for each network is extracted from each test in turn by using a gold standard image as an indicator of both the true edge location and the ideal edge strength at that given position.

3.2 Performance Evaluation

Regardless of the technique used for edge detection; Sobel, Prewitt [1], Canny [2], or statistical [9, 14], the evaluation will be the same and an effective performance measure will be suitable for all. Although this is commonly understood, there has been little development of an effective objective performance measure for any type of detector. Many evaluation methods are subjective in their nature, usually conducted using the human perception of an edge, or with the aim of recognizing the objects within the image. Such subjective evaluation is not regarded as a satisfactory performance criterion and overall an objective, or a combination of both subjective and objective measures [18], is more preferable. The performance of the edge detector used here is assessed on grey-scale images using pre-segmented gold standard images which have been edited by an expert in the field to include only their true edges. A well known performance measure is Pratt's figure of merit (PFOM) [19], this measure is accepted as a common benchmark although it is designed for use with binary images only:

$$R = \frac{1}{I_{sum}} \sum_{i=1}^{I_A} \frac{1}{1 + \beta d_i^2} \tag{4}$$

where $I_{sum} = \max(I, I_A)$, I is the amount of the ideal edge points, I_A is the amount of the detected edge points, and d_i is the distance of the i^{th} edge point from the ideal edge point. β is a scaling constant (typically set to $\frac{1}{9}$). For our purposes we have adapted Pratt's Merit figure (PFOM) to work on grey-scale images (GFOM). The overall performance is calculated for each of the 256 grey levels in both the edge detected and the ideal gold standard images. By thresholding the image at each grey-scale level we can calculate a performance value for that specific level. The product of these 256 performance values is calculated and the mean is assigned as the overall performance value for the image (see Fig. 3).

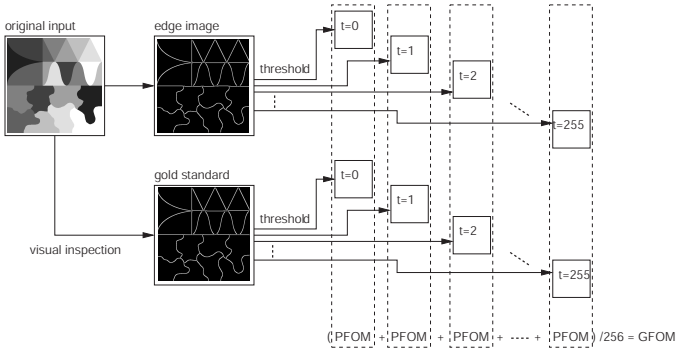


Figure 3. Pratt's FOM adapted to work with 256 grey-scale levels (GFOM). Pratt's figure of merit is calculated in turn for each 256 grey-scale level in both the edge image and the gold standard. The sum FOM is calculated and the mean is allocated as the grey-scale figure of merit (GFOM) for the image.

Image	DSU	TDS	Canny	Image	DSU	TDS	Canny
Mouse Atlas A	0.626	0.628	0.330	SF Bin	0.520	0.516	0.399
Mouse Atlas B	0.579	0.555	0.320	SF Block	0.481	0.553	0.601
Mouse Atlas C	0.538	0.485	0.215	SF Fan	0.518	0.561	0.522
Mouse Atlas D	0.528	0.561	0.272	SF Gloves	0.483	0.564	0.433
Mouse Atlas E	0.618	0.559	0.542	SF Pillow	0.553	0.724	0.368
Mouse Atlas F	0.576	0.580	0.493	SF Tools	0.622	0.497	0.525
mean	0.577	0.561	0.362	mean	0.530	0.570	0.474

Table 1. Grey-scale performance (GFOM) for the multi-statistic detector (mask size 5×5) compared to the Canny filter ($\sigma = 2$). DSU = architecture for Difference of Boxes, Likelihood, and Mann Whitney U tests. TDS = architecture for Student's T, Difference of Boxes, and Likelihood tests. The mean across the range of images is also illustrated. The values are within the range $\langle 0; 1 \rangle$. Value "0" corresponds to poor quality while "1" is an ideal result.

4 Results and Discussion

Several neural network architectures were investigated during the course of this work, with the typical configuration found to be dependent on the number of statistical tests used. This relationship of input neurons to hidden neurons was optimum with twice the number of hidden neurons to input neurons. Table 1 shows the two best combined statistical test architectures applied to the sample images (Fig. 4). The DSU architecture illustrates the results for a network trained on one of each of the three types of statistical tests illustrated here, namely the Difference of Boxes (mean based), Likelihood (variance based), and Mann Whitney U (ranking test). The second architecture (TDS) is a combination of the three tests that each performed well when singly applied to images [17], (the Student's T test, Difference of Boxes, and Likelihood Test). The filter is compared to the Canny filter using a mask size of 5×5 and a comparable Canny $\sigma = 2$. The results clearly illustrate how this novel filter outperforms the traditional Canny filter across a variety of images where edges are both texture (Mouse Atlas), and intensity based interfaces (South Florida images). Furthermore if we look at the mean GFOM across the entire range of images it is clear the Canny filter is consistently outperformed by the multi-test approach. A simple visual inspection of the result images (Fig. 5) would indicate the three detectors appearing comparable across single images. A closer inspection of the Canny output would however further illustrate

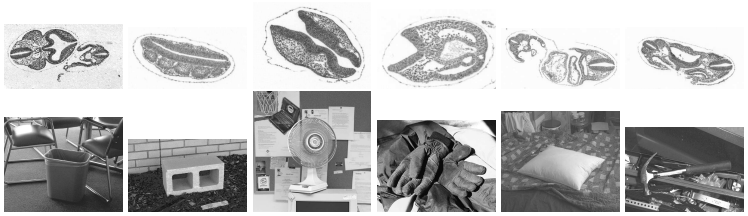


Figure 4. Original Mouse Atlas Images (A–F) – upper row (from left to right). South Florida Images (Bin, Block, Fan, Gloves, Pillow, Tools) – bottom row (from left to right).

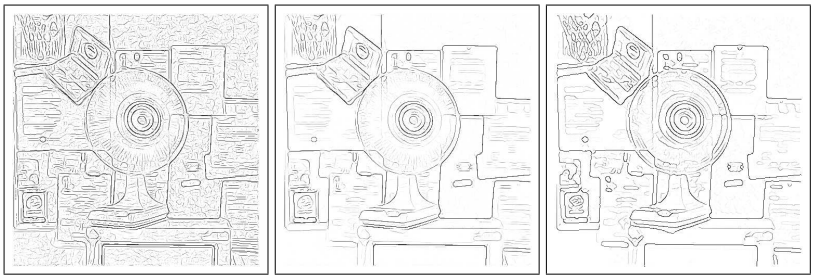


Figure 5. An example of three different results (DSU, TDS and Canny – from left to right) for the South Florida Fan image. The quality is measured using GFOM – see Table 1.

the problems inherent to gaussian smoothing where some required edges are removed, and the corners become rounded and less accurate. This problem illustrates how visual assessment is not always a reliable indication of edge detection quality and objective assessment across the image range gives a more robust performance measure. If we further assess the GFOM of the two multi-test architectures alone we can clearly see their quality is comparable. If evaluated on a larger image set with a greater variety of texture the DSU architecture, although more sensitive, could be expected to outperform the TDS architecture simply by assessing all three types of statistics we have illustrated. Further research will focus on looking for best configuration of NN and tests to provide the optimal edge detector for all images.

Conclusion A novel and successful multi-statistic edge detector has been developed which through the use of neural networks allows different statistical tests to be computed over several mask based regions of an image. This technique can robustly outperform the traditional Canny filter in accurately locating edges across a representative range of images. The accuracy of the detectors was extensively tested using a novel grey-scale performance measure (GFOM) which illustrates a robustness in the results unavailable with visual inspection alone.

Acknowledgement This work was supported by the Ministry of Education of the Czech Republic (Projects No. 1K05021 and No. LC535). We would also like to thank (EMAP) the University of Edinburgh Mouse Atlas project for providing us with the MA images, and also the University of South Florida for the USF image data set.

References

1. M. Šonka, V. Hlaváč & R. Boyle. *Image Processing Analysis and Machine Vision*. Chapman and Hall Publishing, 1986.
2. J. Canny. "A computational approach to edge detection." *IEEE T-PAMI* **8**, pp. 769–698, 1986.
3. P. Perona & J. Malik. "Scale-space and edge detection using anisotropic diffusion." *IEEE T-PAMI* **12(7)**, pp. 629–639, 1990.
4. P. de Souza. "Edge detection using sliding statistical tests." *CVGIP* **23(1)**, pp. 1–14, 1983.
5. A. C. Bovik, T. S. Huang & D. C. M. Jr. "Nonparametric tests for edge detection in noise." *Pattern Recognition* **19(3)**, pp. 209–219, 1986.
6. M. Beauchemin, K. P. B. Thomson & G. Edwards. "On nonparametric edge detection in multilook sar images." *IEEE Transactions on GRS* **36(5)**, pp. 1826–1829, 1998.
7. J. S. Huang & D. H. Tseng. "Statistical theory of edge detection." *CVGIP* **34(3)**, pp. 337–346, 1988.
8. E. Guest. *Automatic reconstruction from serial sections*. Ph.D. thesis, Department of Artificial Intelligence, University of Edinburgh, 1994.
9. M. N. Fesharaki & G. R. Hellestrand. "A new edge detection algorithm based on a statistical approach." In *ICSIPNN '94*, pp. 21–24. 1994. IEEE.
10. A. Kundu. "Robust edge detection." *Pattern Recognition* **23(5)**, pp. 423–440, 1990.
11. Z. Hou. "Robust edge detection." *Pattern Recognition* **36(9)**, pp. 2083–2091, 2003.
12. D. Lim & S. Jan. "Comparison of two-sample tests for edge detection in noisy images." *Statistician* **51(1)**, pp. 21–30, 2002.
13. D. Lim & S. Jan. "Robust edge detection in noisy images." *Comp. Statistics and Data Analysis* **50(3)**, pp. 803–812, 2006.
14. N. J. Bowring, E. Guest, P. Twigg et al. "A new statistical method for edge detection on textured and cluttered images." In 4^{th} *IASTED VIIP Conf.*, pp. 435–440. 2004.
15. I. Williams, N. J. Bowring, E. Guest et al. "A combined statistical/neural network multi-scale edge detector." In 5^{th} *IASTED VIIP Conf.* 2005. ISBN: 0-88986-528-0, ref: 480-266.
16. A. Rosenfeld. "A nonlinear edge detection technique." *Proceedings of the IEEE.* **58(5)**, pp. 814–816, 1970.
17. D. Svoboda, I. Williams, N. Bowring et al. "Statistical techniques for edge detection in histological images." In 1^{st} *Int. Conf. on VISAPP*, pp. 457–462. INSTICC Press – Portugal, 2006. ISBN 972-8865-40-6.
18. M. Heath, S. Sarkar, T. Sanocki et al. "Comparison of edge detectors: A methodology and initial study." *CVIU* **69(1)**, pp. 38–54, 1996.
19. E. Abdou, I & K. Pratt, W. "Quantitative design and evaluation of enhancement/thresholding edge detectors." *Proceedings of the IEEE.* **27(5)**, pp. 753–763, 1979.

A Virtual Ultrasound Imaging System for the Simulation of Ultrasound-Guided Needle Insertion Procedures

Yanong Zhu[a*], Derek Magee[a†], Rish Ratnalingam[b] and David Kessel[b]

[a] School of Computing, University of Leeds, Leeds, UK
[b] Leeds Teaching Hospitals NHS Trust, Leeds, UK

Abstract. In this paper we describe a virtual ultrasound imaging system for the simulation of ultrasound guided needle insertion procedures, which is designed to improve the early training stages of interventional radiology trainees. A pair of calibrated magnetic 3D motion sensors are used to capture the position and orientation of a mock ultrasound probe and needle, whilst emulational ultrasound images are generated in real-time from a labelled volumetric data set that is non-rigidly aligned to a physical model of human body. To achieve a realistic simulation of ultrasound imaging, a set of volumetric textures are constructed to represent the typical appearance of ultrasound images, and an alpha blending method is applied to produce the radial blurring effect. The procedures of volumetric registration, sensor calibration, construction of texture bank, and image rendering, are presented.

1 Introduction

The insertion of a needle (or equivalent device) into a patient is frequently performed in a variety of clinical routines, such as drainage of abscess, relief of blockages in the kidneys or the liver, radioactive seed implantation, and biopsy of deep tissues. To ensure an accurate placement of the needle, such procedure is usually guided by real-time feedback from a 2D ultrasound display. However, acquisition of the spatial reasoning and hand-eye co-ordination skills for such ultrasound-guided needle insertion procedure requires a specialised training process. Conventionally, this training is performed on the basis of the so-called *see one, do one, teach one* paradigm [1], in which the trainees, after observing a large number of such procedures, practice on live patients under supervision of skilled radiologists. The evaluation of the acquired skills in such training process is often subjective, and it is often risky for the patients when such invasive procedure is performed by less skilled radiologists.

An obvious solution is using simulation systems for the training of such interventional procedures. Some systems based on physical phantoms have been developed (e.g., the breast ultrasound phantom from SIEL Ltd.). However, such physical systems are often designed for specific abnormalities, and high in cost because of the use of a real ultrasound machine. A more feasible alternative is computerised simulation, based on virtual reality (VR), in which the complete environment is simulated within a computer, or augmented reality (AR), in which physical and virtual worlds are interconnected. Integrating computer visualisation techniques and physical components such as haptic feedback devices and/or mannequins, AR simulators can be used to achieve a more realistic experience for the training of interventional operations with visual feedbacks. Holbrey $et\ al$ [2] present a simulation system for vascular suturing training, based on a haptic feedback device, stereo display, and Finite Element Models. Gorman $et\ al$ [3] developed a system incorporating a mannequin and a haptic feedback device for the simulation of lumbar puncture procedure, which is based on offline CT data, and so has no real-time visual feedback. Simulators for fluoroscopy (real-time or near real-time X-ray based 2D imaging) guided procedures [1, 4] are closer in nature to the ultrasound guided simulation system presented in this paper. The advantage of fluoroscopy over ultrasound is that the direction of imaging is far less important than with ultrasound due to the absorbency/reflection properties of ultrasound. However, ultrasound is fully real-time, flexible, and inherently safe, while exposure to X-ray should be minimised. Alterovitz $et\ al$ present a simulator for ultrasound guided needle insertion [5]. However, this is a 2D computer only simulator, which does not simulate probe motion or any 3D aspects of the procedure.

We have previously described a prototype simulation system for ultrasound guided needle insertion procedures [6]. In this initial system the physical model of the human body is a plastic mannequin, on which the user can perform operations such as ultrasound scanning and needle insertion. The position and orientation of the mock ultrasound probe and the needle are captured by a pair of magnetic 3D motion sensors. Virtual ultrasound images are generated from a volumetric CT data set that is non-rigidly registered to the mannequin surface, and rendered in real-time to guide the needle insertion procedure. The initial evaluation of the system, performed using a set of unskilled individuals and a set of experienced clinicians, reveals that, after structured practice on the simulator,

*Email:yanong@comp.leeds.ac.uk
†Email:drm@comp.leeds.ac.uk

the positioning and hand-eye co-ordination abilities can be clearly improved. However, this initial system has a number of shortcomings. Firstly, since the synthetic images are generated directly from a volume of CT data, they look more like CT images, rather than the desired ultrasound feedback. Additionally, the plastic mannequin surface is too hard for a needle to penetrate. Hence instead of a real needle that can be used to insert into the body, the sensor is attached to a wooden stick via a sliding mechanism to mimic the insertion process. The system has been further developed to achieve more realistic simulation of the procedure. In addition to the construction and use of a penetrable model of human body, a Laplacian pyramid based approach is developed to generate synthetic textures from ultrasound image samples, and a 2D alpha blending technique is adopted to produce the ultrasound-specific radial blurring effects. With the emphasis on these newer improvements, this paper also briefly describes other aspects of the system (e.g. volume registration and sensor calibration) to provide a complete overview of the system.

2 System Implementation

The hardware of our simulation system consists of three components, a standard PC (Dual Intel® Xeon™ CPU 3.20 GHz, 2.0 GB of RAM, running Microsoft® Windows® XP™ Professional), a full scale penetrable model made of latex plastic and foam, and a pair of Ascension PCIBirds magnetic 3D position/orientation sensors. One of the sensors is rigidly attached to a spare ultrasound probe, and the other attached to a standard biopsy needle. Figure 1 illustrates the hardware setup of our simulator.

To produce synthetic ultrasound images, there are a number of tasks to be accomplished (Figure 2). Firstly, a virtual human body, in the form of a volumetric CT data set, is obtained from a human subject. The volume is manually segmented so that each voxel in the volume is assigned a label that indicates the property of the voxel, and then registered to the surface of the physical model (Section 2.1). Secondly, the motion sensors are calibrated so that the position and orientation of the probe/needle they attached to can be correctly captured and represented (Section 2.2). Thirdly, a bank of volumetric textures are constructed from real ultrasound images to represent the typical appearance of the objects in the segmented volumetric data (Section 2.3.1). During interaction, given the position/orientation of the probe, a scan plane is determined. The image is produced by raster scanning through the pixels in the ultrasound portion of the plane, and projecting each point to the corresponding point in 3D space so that its greylevel can be determined. In addition, a 2D ray-casting method is applied to produce the shadow effects caused by bones and air bags in the virtual body. The constructed image is then rendered using an alpha blending routine to simulate the radial blur effect (Section 2.3.2).

Figure 1. Overview of hardware setup.

2.1 Registration of Volumetric Data to The Model Surface

The simulator renders "virtual ultrasound images" based on a CT volumetric data set acquired from a real human. The data is first manually segmented by our cooperating expert radiologist, then registered to the surface of the physical model in a three-stage process. Firstly, the body surface embedded in the volumetric data is extract by manually annotation of the slices, and the model surface is extracted by passing a calibrated motion sensor over the surface of the mannequin model. The second stage involves a rigid alignment of the surfaces by the Iterative Closest Point (ICP) algorithm [7]. Subsequently, for the surfaces to be completely aligned, a non-rigid registration process, based on the RanSaC algorithm [8] is performed. This yields a quadratic function, which can be used to warp the volumetric data offline to a new volume that is aligned to the model surface. Once the warped volume has been formed, online calibration is simply a process of rigid scale and offset between the volume co-ordinate system and the world co-ordinate system.

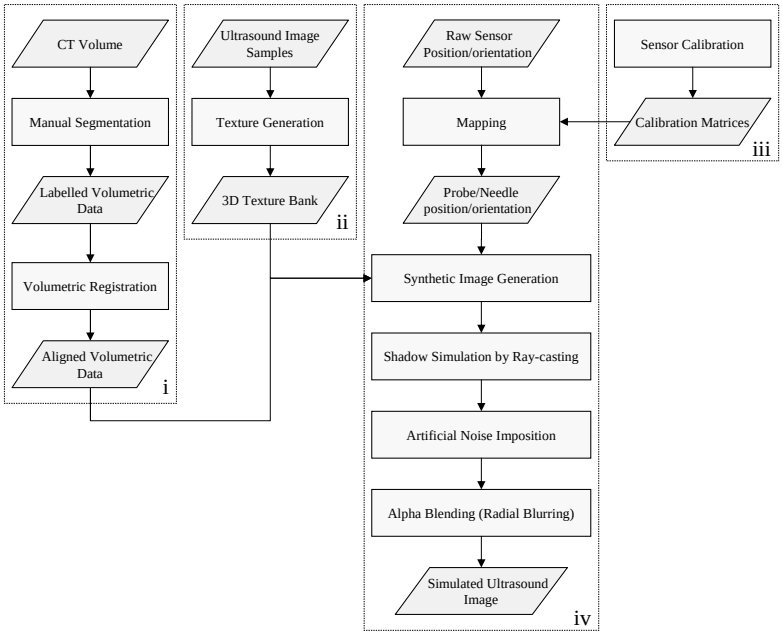


Figure 2. Simulation process of ultrasound imaging.

2.2 Sensor Calibration

Although it is not always possible to place the motion sensor at the position that we want to measure (e.g. the tip of the needle), or orient it in the direction we want to measure, it is possible to rigidly attach the sensors to the probe and needle, and estimate a calibration function to map the sensor output to the desired values. In case of the ultrasound probe, the calibration point is located at the centre of the end of the probe, and the mapping function from the attached sensor to the calibration point is estimated from multiple unique position samples using a standard Least Square Fitting method. Similar approach can be applied to the calibration of the probe direction, as well as the calibration of the needle. A complete description of the sensor calibration process can be found in [6].

2.3 Generation of Synthetic Ultrasound Images

Ultrasound guided needle insertion requires real-time visual feedback from the simulator. It is difficult to model the physical properties of an real ultrasound system. However, it is possible to simulate these properties by a series of common 2D effects. These include texture mapping, artificial noise imposition, and radial blurring.

2.3.1 Construction of Texture Database

A set of volumetric "solid textures" [9, 10] (one for each visually distinct tissue type) is learned from sets of hand selected 2D ultrasound examples. Solid textures are volumes for which the surfaces appear similarly textured if cut by any plane (or other surface). There is a large body of literature on the generation of synthetic 2D textures, but much less on solid texture generation. This is probably due to the fact that they are not currently as widely used for a number of reasons including; Consumer graphics hardware architectures, higher memory requirements than 2D textures, and the comparative difficulty in making volume textures anisotropic. However, they are ideally suited to our application as 2D textures cannot model the effect of sensor motion in three dimensions. 2D synthetic texture methods fall broadly into two categories; i) Cut-and-paste based methods (e.g. [11]), in which patches of training images are pasted together to form an new texture, and ii) Image statistics based methods (e.g. [12]). We have found approach i) does not easily extend to learning 3D solid textures from 2D examples. Our approach is to use a Laplacian image pyramid [12] (with 3 or 4 levels) to represent the image and volume statistics, and to match the statistics of the volume to the training images. A Laplacian image pyramid is a set of images (or volumes) generated from a single image by iteratively applying a low pass filter and subtracting the result from the previous image generated. This is a form of frequency decomposition. We use a 256 bin greylevel histogram at each level of the Laplacian pyramid to capture the training data statistics ($P_l(g)$). We also form two dimensional histograms

to capture the joint greylevel probability distribution of each level and the level below ($P_{l,l-1}(g_l, g_{l-1})$). To generate the solid texture we start at the lowest (most coarse) level of the Laplacian pyramid and sample greylevel values from $P_l(g)$. Subsequent levels of the pyramid are generated based on a weighted combination of $P_l(g)$ and $P_{l,l-1}(g_l, g_{l-1})$:

$$P_l(g) = K_l P_l(g) + (1 - K_l) P_{l,l-1}(g_l, g_{l-1}) \quad (1)$$

Where K_l is typically 0.6 for all levels except the top (highest frequency) level, where it is 0. The use of $P_l(g)$ compensates for the data sparsity at the lower levels which can lead to textures that are too regular. This is unnecessary at the top (highest frequency) level as sufficient data is available. A tessellating solid texture (typically 32x32x32) is generated from the Laplacian volume pyramid generated by the sampling process by use of a low-pass filter that assumes 3D wrapping at the edges of the volume. This is a very fast process as it does not rely on searching through the training set (e.g. [13]) or iterative optimisation (e.g. [12]). Examples of training textures and slices through generated solid textures are given in figure 3.

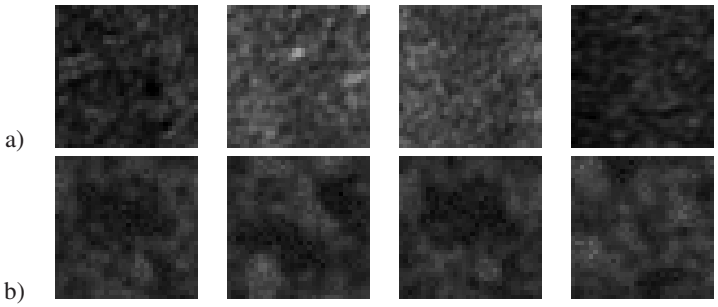
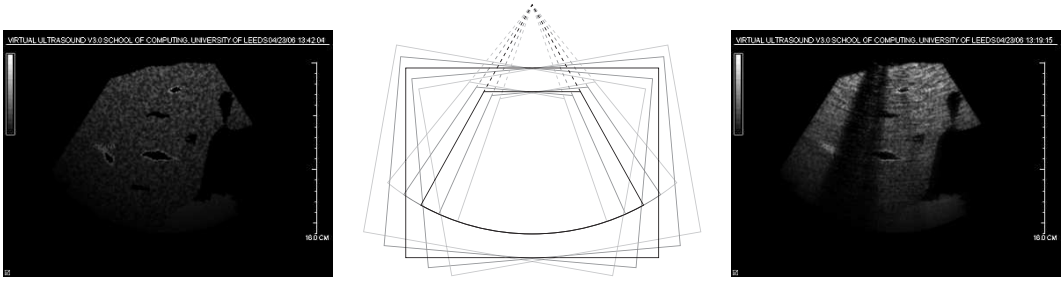


Figure 3. a) Training textures (Liver), b) Slices through generated texture

2.3.2 Simulation of Ultrasound-Specific Artifacts

With the generated textures, an initial ultrasound image can be generated by raster scanning through the pixels in the ultrasound portion of the scan plane, and projecting each pixel to the corresponding voxel in the 3D volume space so that its label can be determined. The pixel's label and its 3D coordinates are then used to sample the corresponding texture to determine the greylevel of the pixel. An example of the intermediate result at this stage is shown in Figure 4 (a). However this image is not much close to real ultrasound output due to the lack of some ultrasound-specific artifacts, such as speckle noise, shadows, and radial motion blurring. The speckle effect can be simply simulated by adding Gaussian distributed artificial noise to the image pixels (We currently use $\sigma = 15$ for the Gaussian noise). Since ultrasound pulses do not penetrate bones or air, a shadow will be cast behind these objects if they present in the scan plane. We simulate the shadowing effect by a 2D ray-casting approach. Regarding the centre of the end of the ultrasound probe as a spot light, we compute the greylevel decrement of each image pixel according to the number of bone pixels (or air pixels) along the straight line segment that connects the spot light and the pixel; the more bone pixels appear along the line segment, the darker the pixel is. In fact, note that the ultrasound image is generated by a top-bottom raster scanning process, both the speckle and shadowing effects can be produced in one scan of the full image, and hence maintain a satisfactory frame rate in real-time interactive mode.

The radial blur effect creates blurs around a specific point in an image, simulating the effects of a swirling camera. This effect can be used to simulate the radial scanning motion of a real ultrasound transducer. However, the traditional method that computes the blurred image by a convolution operation is too computationally expensive to be applied in real-time. Therefore, we have made use of the power of modern graphics cards to achieve the same effect. The traditional blurring convolution can be viewed as a weighted sum process that adds up the original image and a number of its weighted copies with certain offsets (or rotations). This can be performed by the alpha blending function that is efficiently implemented by most modern graphics cards.We draw the generated image many times to create the radial blur effect, varying the angle of rotation about the blurring centre, and increasing the alpha level while the angle increases. Figure 4 (b) is an illustration of the alpha blending process, and an example of the final ultrasound image generated by our simulator is given in Figure 4 (c). It is worth to emphasise that, though consisting of a series of tasks, the image generation and rendering process is still much efficient in computation. The average frame rate of our ultrasound simulator is 12-13 fps, which is very close to the 16 fps rate that is commonly used in clinical radiology.



(a) (b) (c)

Figure 4. (a) A "raw" synthetic ultrasound image without additional artifacts; (b) Illustration of the alpha blending process for radial blur effect; (c) An example of the final simulated ultrasound image.

3 Discussion

We have presented the latest developments of our simulation system for ultrasound guided needle insertion procedures. These developments have largely improved the fidelity of the simulator, in terms of the synthesis of much more realistic ultrasound images, and a better model of the human body on which needle insertion can be performed. Our collaborative radiologists were much impressed by the simulated procedure and its reality was highly approved, though an objective assessment of the system is still indispensable. Future work will be focused on three aspects. Firstly, at present only a model of the front body is in use, which means that simulation of needle insertion at the back of the body (e.g. needle insertion into the kidneys) is not currently available. A model of the back human body is being created, and shortly it will be added to the system. Secondly, the current simulator is based on a static volumetric data set. Therefore motions of internal organs caused by patient breathing and heart beating, which can be observed in real ultrasound imaging, are not modelled. A long-term plan is to construct a self-deforming model of the human body, and thus simulate the motions of the internal anatomies. Thirdly, based on the initial evaluation of the system as presented in [6], a comprehensive evaluation scheme, covering clinical radiology, psychology and computer sciences is being designed and will be performed in due course.

References

1. Y. Wang, C. Chui, H. Lim et al. "Real-time interactive simulator for percutaneous coronary revascularization procedures." *Computer Aided Surgery* **3(5)**, pp. 211–227, 1998.
2. R. Holbrey, A. Bulpitt, K. Brodlie et al. "A model for virtual suturing in vascular surgery." In *Proc. Theory and Practice of Computer Graphics*, pp. 50–58. IEEE Computer Society Press, 2004.
3. P. Gorman, T. Krummel, R. Webster et al. "A prototype haptic lumbar puncture simulator." In *Proc. Medicine Meets Virtual Reality*, pp. 106–109. 2000.
4. Z. Li, C. Chui, J. Anderson et al. "Computer environment for interventional neuroradiology procedures." *Simulation and Gaming* **32(3)**, pp. 405–420, 2001.
5. R. Alterovitz, J. Pouliot, R. Taschereau et al. "Simulating needle insertion and radioactive seed implantation for prostate brachytherapy." In *Proc. Medicine Meets Virtual Reality*, pp. 19–25. 2003.
6. D. Magee & D. Kessel. "A computer based simulator for ultrasound guided needle insertion procedures." In *Proc. IEE International Conference on Visual Information Engineering*, pp. 301–308. 2005.
7. P. Besl & N. McKay. "A method for registration of 3D shapes." *IEEE Trans. on Pattern Analysis and Machine Intelligence* **14(2)**, pp. 239–256, 1992.
8. M. Fischler & R. Bolles. "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography." *Communications of the Association for Computing Machinery* **24(6)**, pp. 381–395, 1981.
9. K. Perlin. "An image synthesizer." In *Proc. International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 287–296. 1985.
10. D. Peachey. "Solid texturing of complex objects." In *Proc. International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 279–286. 1985.
11. A. Efros & W. Freeman. "Image quilting for texture synthesis and transfer." In *Proc. International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 341–346. 2001.
12. D. Heeger & J. Bergen. "Pyramid-based texture analysis/synthesis." In *Proc. International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 229–238. 1995.
13. J. DeBonet. "Multiresolution sampling procedure for analysis and synthesis of texture images." In *Proc. International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 361–368. 1997.

Analysis of multispectral images of the colon to reveal histological changes characteristic of cancer

Džena Hidović-Rowe[a], Ela Claridge[a], Tariq Ismail[b] and Phillipe Taniere[b]

[a]School of Computer Science, The University of Birmingham
[b]Birmingham University Hospital, Birmingham
{D.Hidovic, E.Claridge}@cs.bham.ac.uk

Abstract: Colon cancer alters the macroarchitecture of the colon tissue. Common changes include the formation of new blood vessels (angiogenesis) and the distortion of the tissue collagen matrix. Such changes affect the colon colouration. This paper presents the principles of a novel image analysis method capable of extracting parameters depicting histological quantities of the colon from its multispectral images. The method is based on a computational, physics-based model of light interaction with tissue. The colon structure is represented by three layers: mucosa, submucosa and muscle layer, and parametrised by the concentration of blood haemoglobins; the size and density of collagen fibres; and the thickness of the layers. Using the entire histologically plausible ranges for these parameters, a cross-reference is created computationally between the histological quantities and the spectra which can be obtained at each point of a multispectral image set. In a pilot study the method was applied to two excised colon samples comprising normal tissue and adenocarcinomas. The histological parameters extracted from the cancerous regions showed the changes characteristic of the colon cancers, namely the increase in the blood volume fraction, decrease in the collagen density and increase in the thickness of the mucosal layer.

1. Introduction

Development of colon cancer alters the macroarchitecture of the colon tissue. The main changes include an increase in microvascularisation and hence the blood content of the tissue [1], and distortion of its collagen matrix [2]. Given that blood and collagen act, respectively, as strong absorbers and scatterers of light in the visible range of the spectrum, those changes alter the colon colouration. At the early stages of the disease the variations in the colour between normal and abnormal tissue tend to be very subtle, and hence not easily discernible by the human eye. Standard red, green, blue (RGB) primaries used in colour imaging provide a fairly impoverished representation of the tissue colouration, and it is likely that spectra arising from different tissue histologies may produce the same RGB values [3] – a phenomenon known as metamerism. This may result in a loss of diagnostically important information. Therefore, a richer spectral data set might be more appropriate for representing subtle changes in colouration. Multispectral imaging combines imaging and spectroscopy, and enables the extraction of the whole spectral content at every pixel of the image, so providing the combined advantages of high spatial and spectral resolutions.

Raw multispectral images are difficult to interpret directly. In remote sensing, which has a long history of using multispectral imagery, image regions are classified into various terrain types based on their spectral signatures. In this paper we present a different approach, where multispectral data is interpreted in terms of the underlying tissue histology. Such interpretation is possible through the use of a physics-based model of light interaction with colon tissue. The result is a set of parametric maps which show the spatial distribution, relative quantities of blood and the density and thickness of the collagen matrix in the topmost layer of the colon. We believe that, in the long term, the explicit presentation of these histological should lead to more reliable and possibly earlier diagnosis.

Our approach differs from traditional and more common approaches to image analysis which are based on statistical analysis of the image features. Instead of classifying tissues as normal or neoplastic, we are concerned with understanding the causes of the changes in tissue spectra and hence colour, and correlating those changes to the histological alterations of the tissue. At the heart of our method is a computational model of colon reflectance. Using the Monte Carlo method of simulating light-tissue interactions, the model correlates tissue parameters which describe its optical properties with the spectral composition of the light that can be observed at the tissue surface after penetrating the tissue and interacting with its structure. The spectra obtained at each point of a multispectral image are then interpreted by the inversion of the model.

This paper presents preliminary results of interpretation of multispectral images of the colon using the above approach. Section 2 presents a method of constructing a forward model of colon. A method of extracting histological parameters from the spectra, and its validation, are presented in section 3. Section 4 is concerned with image acquisition and post-processing of multispectral images in order to extract the spectral content at each pixel. Results, in the form of parametric maps, and their analysis are presented in section 5.

2. Forward modelling of colon reflectance: from tissue parameters to spectra

We have developed a computational, physics-based model of light interaction with tissue, which predicts reflectance spectra associated with specific instances of colon tissue. Here we give just a brief description of the model which is described in more detail in our previous paper [4].

The colon structure is represented by three layers: mucosa, submucosa and smooth muscle, which interact with the light incident on the surface of the colon. In all colon layers, the strongest absorbers of light are blood haemoglobins, while scattering is mainly due to collagen and subcelullar organelles. These components are characterised by different properties and content in colon layers, which as a consequence have different optical properties, and cause different light-tissue interactions. Our analysis have shown that the parameters characterizing the first layer, the mucosa, have the most significant effect on behaviour of the remitted spectra, whereas the deeper layers determine the relative magnitude of the spectra and produce small changes in the red end of the spectrum [4]. Therefore the colon parametrisation includes variable parameters (N=5) characterizing the mucosal layer only: blood volume fraction, haemoglobin saturation (the ratio of oxygenated and de-oxygenated blood), the size of collagen fibres, their density and the layer thickness. The parameters characterizing the submucosa and the muscle have been assumed to have constant values, which have been set to the median of their relative ranges. Although this restricts the analysis to the mucosal tissue, it should be still of diagnostic value, as the early changes of the colon normally happen in this layer. For detailed description of parameters and the sources of the parameter values see [4].

Once the histological parameters and their normal ranges of values are defined, it is possible to create a cross-reference between these parameters and the reflectance spectra, and subsequently, a correspondence between the histology and images. The tissue reflectance is defined by N (=5) variable parameters. Each *specific instance* of tissue can thus be defined by an N-dimensional parameter vector $\mathbf{p} = [p_1 \ldots p_n]$. The range of each parameter is discretised to k_n levels, giving in total $K = k_1 \times k_2 \times k_n$ parameter vectors which, together, define *all the possible instances* of the colon tissue. Each parameter vector **p** can be associated with a spectrum by simulating light propagation in the tissue depicted by **p** using one of the methods for approximating the solution to the light transport equation, such as for example Monte Carlo (MC) [5]. Subsequently, multispectral image values can be obtained from the computed spectra by convolving each spectrum with a set of appropriate bandpass filters, f_m, m=1,…,M. Through such modelling of the light interaction with tissue and of the image acquisition process, we associate with each parameter a spectrum, and also an M-dimensional image vector $\mathbf{i} = [i_1 \ \ldots \ i_M]$. All the parameter vectors **p** together with the respective image vectors **i** form the *tissue reflectance model*. This model is used to derive parameters from multispectral images of tissue (see section 4).

Reflectance spectra of the colon tissue were calculated for a discrete set of the (K=) 20 wavelengths chosen to best describe the colon spectra: {450, 480, 506, 514, 522, 540, 548, 560, 564, 568, 574, 580, 586, 594, 610, 620, 630, 640, 676, 700}. Each spectrum was obtained by MC simulation using 200,000 photons. The modelled spectra were validated by comparison with the experimental data acquired *in vivo* using a fibre optics probe during colonoscopy procedures [6]. If the model is correct, it must be able to produce the spectra obtained by measurements (this, however, is just a necessary, but not also a sufficient condition). The validation data comprised 50 spectra from histologically confirmed non-neoplastic tissue (normal mucosa) and 7 spectra from cancerous tissue. Parameter values extracted from cancerous tissue were compared to the values extracted from normal colon tissue of the same patient. The method of parameter extraction and the validation results are described in the next section.

3. Extraction of parameters from spectral data

Extraction of the parameter values from the spectra was implemented as an optimisation procedure. The goal was to find a set of parameter values such that a spectrum generated from the parameters using the Monte Carlo simulation provides the best match to the measured spectrum. The criterion for the best match is the minimum distance between the two spectra defined as the mean absolute difference between the spectral values at corresponding wavelengths. The optimisation was implemented using an evolutionary strategy [7] which can be briefly summarised as follows. At each step, a set of parameters is chosen and the corresponding spectrum calculated. If the distance between the modelled and measured spectrum is smaller than some fixed threshold, the procedure is terminated. Otherwise, a new set of parameters is chosen according to a continuous version of the Gray-code neighbourhood distribution [7] and the process repeated.

This procedure was applied to the 57 experimentally obtained colon spectra described in section 2. In all the cases the optimisation procedure found a well matching spectrum in the reflectance model (figure 1). The matching error, measured as the mean absolute distance between the spectra, was 0.01 ± 0.0019. Parameters

extracted from the measured spectra were well within the histologically plausible ranges. Encouragingly, parameters extracted from the spectra of the abnormal tissue were characterised by increased blood content and decreased collagen density in comparison to the normal tissue of the same patient [8]. This is consistent with known histological differences between normal and abnormal colon tissue.

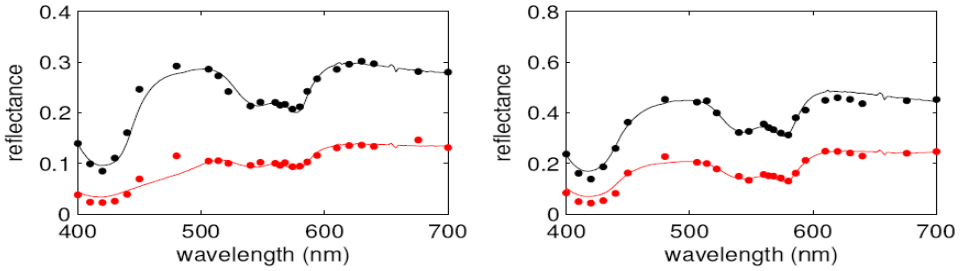


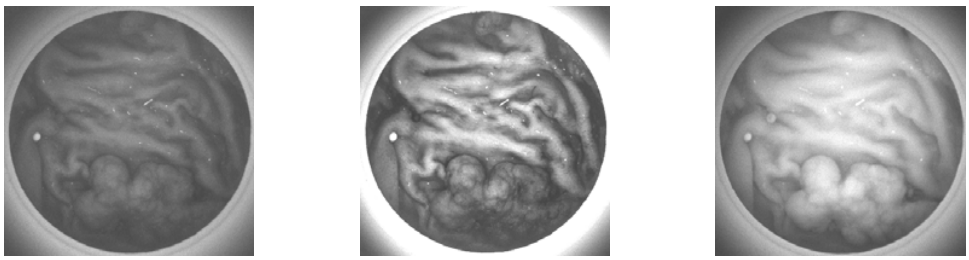
Figure 1. Model verification. Two examples of the measured spectra and the corresponding model spectra. Measured spectra are shown as solid lines, and modelled spectra as filled circles. Spectra of the normal tissue are shown in black, and of the cancerous tissue from the same patient are shown in red (grey in a monochrome reproduction).

4. Extraction of parameters from multispectral images of the colon

Following validation on the spectral data, we have carried out a preliminary experiment on the data obtained from multispectral images of *ex-vivo* colon tissue. *Ex-vivo* samples were used as currently there is no established endoscopic technology for multispectral imaging of the intact colon. The objective of the experiment was to investigate whether the parametric maps obtained from the multispectral image data show the spatial distribution and the magnitudes of the individual histological parameters consistent with the histology of a given sample.

4.1. Imaging

In a pilot study, images of two tissue samples were acquired within one hour of the surgical extraction performed at the Queen Elizabeth Hospital, Birmingham. In each sample, images were obtained for the apparently healthy tissue, at a distance more than 2 cm away from the histologically confirmed cancerous tissue (not shown), and the cancerous site itself (Fig. 2). The system used for acquiring the images consisted of a Retiga EXi (QImaging, Canada) 12bit monochrome camera, VariSpec (CRI, U.S.A.) liquid crystal tunable filters, and an Integrating Sphere (ProLite, U.K.) which produces a smooth spatial and spectral illumination. The VariSpec filters allow the selection of Gaussian-shaped filters of halfwidth 5 nm in the range from 400 to 700 nm. The images were acquired through a 3 inch aperture of the sphere. The total acquisition time per sample was under 5 seconds. An example set of images of the colon including cancerous tissue (adenocarcinoma) is shown in Figure 2. To ensure the spatial correspondence, the images in each set were registered.



455 nm 540 nm 676 nm

Figure 2. Images of the cancerous colon tissue at wavelengths 455, 540 and 676 nm. Blood vessels and smaller features of the tissue can be seen in the image 540 nm where blood absorption is strong. At 455 nm the features are less distinct because high blood absorption coefficient causes even small amounts of blood to absorb strongly thus reducing differentiation between high and low blood levels. At 676 nm those features seem again to disappear because of low absorption of light by haemoglobin derivatives present in the blood and increased scatter by collagen, cells and organelles in the mucosal layer.

4.2. Extraction of the spectra and construction of the parametric maps

Each pixel *(x,y)* in a multispectral image set can be represented by an image vector $\mathbf{i} = [i_1 \ldots i_M]$ where the *m*-th component of the vector is the image value $i_m(x,y)$ corresponding to *m*-th spectral band. This value is a function of the tissue properties, the incident light and the characteristics of the image acquisition system:

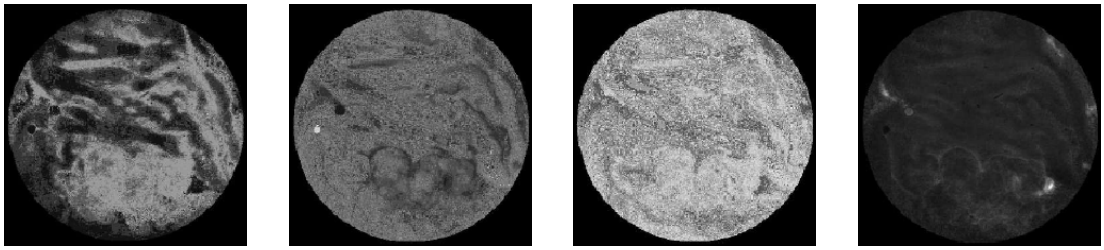
$$i_m(x,y) = \int_{\lambda} R(x,y,\lambda) \cdot I_0(x,y,\lambda) \cdot q(\lambda) \cdot f_m(\lambda)\, d\lambda \quad (1)$$

where $R(x,y,\lambda)$ *is* tissue reflectance at the pixel (x,y) at the wavelength λ, $I_0(x,y,\lambda)$ is intensity of incident light at (x,y) and wavelength λ, $q(\lambda)$ is the quantum efficiency of the camera at λ, and $f_m(\lambda)$ is the spectral response of the m-th filter at the wavelength λ. It can be seen that the image vector **i** depends on the spectral reflectance of the tissue, $R(x,y,\lambda)$, as well as on the parameters characterising the imaging system. By deconvolving $i_m(x,y)$ with $I_0(x,y,\lambda)$, $q(\lambda)$ and $f_m(\lambda)$, which are available as a function of wavelength from instrument calibration, it is possible to obtain the "pure" tissue reflectance spectrum at each point of the multispectral image.

Using the parameter extraction method described in section 3 above, the histological parameters were derived from the tissue reflectance spectra $R(x,y,\lambda)$ at each pixel, and their magnitudes were recorded in the five parametric maps representing the mucosal blood volume fraction, haemoglobin saturation, the size of collagen fibres, their density and the thickness of the mucosal layer. In addition to these parameters, two additional useful quantities were stored in the parametric maps: the scale factor, which is a function of the distance between the camera sensor and the imaged tissue (and reflects the surface topography); and the match error, which captures the degree of discrepancy between the modelled and the measured spectrum. This latter may have a diagnostic significance, as discussed below.

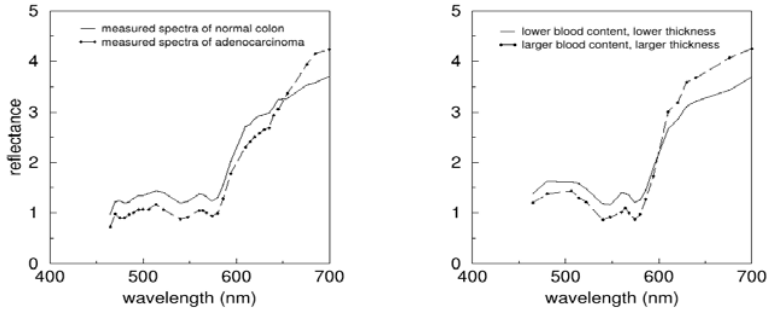
5. Results

Figure 3 shows three out of five parametric maps of a colon tissue sample shown in figure 2 (the other two maps did not show significant variations). A tumour, histologically confirmed as an adenocarcinoma, is clearly visible in the lower part of the image. The maps show the increased blood volume fraction, the decrease in the collagen density and increase in the mucosal thickness in the region of the cancer, as compared to the tissue further away from the tumour. This is in agreement with known histological changes which occur with development of the cancer. The error map shows a slightly increased error between the measured spectra and the spectra found in the reflectance model. This could indicate that the tissue in the cancerous region is distorted beyond the limits represented in the model, and as such, may provide an additional sign of abnormality.



(a) (b) (c) (d)

Figure 3. Parametric maps derived from a multispectral image set illustrated in Fig. 2. The maps show (a) blood volume fraction; (b) collagen density; and (c) mucosal layer thickness (d) Error map. In all the maps brighter = a higher value.



(a) (b)

Figure 4. Comparison of spectra of normal and cancerous colon tissue from figure 3. (a) The measured spectra: solid line – normal tissue; dashed line - adenocarcinoma. (b) Analogous spectral differences are visible in the modelled spectra if blood content and thickness are increased (dashed line) with respect to normal tissue (solid line).

A comparison of the spectra of cancerous and normal tissue is shown in figure 4. The main spectral variations seem to be due to larger blood content of the cancerous growth (corresponding spectrum is lower in blue and green bands), and larger mucosal thickness (the red end of the spectrum is higher). The spectra are consistent

with known histological variations that occur with development of malignancy. Analogous spectral variations were reproduced by our model, and shown in the plot in figure 4(b).

6. Conclusions and future work

The model correctly predicts the spectra, and thus colours, of colon tissue, indicating that histological quantities can be computed from multispectral images acquired by a digital camera equipped with suitable filters. The model validation [4] suggested that the model should be capable of differentiating between normal and abnormal tissue, which would potentially make it a useful non-invasive aid in the early detection of colon cancer. A preliminary study described in this paper has extended the parameter extraction method from the pure spectra [4] to the multispectral images of the colon. These very first results obtained for *ex-vivo* colon samples indicate that parametric maps of the blood volume fraction, the collagen density and thickness of the mucosa show the differences between the normal tissue and the adenocarcinoma which are consistent with the colon histology.

Whereas the forward model has been satisfactorily validated [4], we do not yet have such evidence for the parameter recovery method. The results described in this paper are promising, but they need to be supported by results from a much larger data set. This is the subject of the ongoing work. Work is also being undertaken in order to optimise the choice of filters to use in the acquisition of multi-spectral images of tissue. Earlier work by our group has shown that under certain conditions it is possible to recover parameter values relatively accurately from a small number of optimally selected spectral bands [9]. Four parameters characterising the mucosa (thickness, blood volume fraction, size, and density of collagen - see section 2) require four bandpass filters, making the image acquisition with the existing endoscope technology feasible.

The next major step will be the acquisition and the interpretation of the multispectral colon images *in-vivo*. The changes to the model will be trivial, as only the parameter ranges will need to be suitably altered to reflect the changes in histology. However, the recovery of the true tissue reflectance spectra ($R(x,y,\lambda)$ in Eq. 1) under the unknown illumination conditions (magnitude of the incident light, its angular distribution, etc.) pose a major challenge. One possible approach has already been developed by our group [10], but it will need to be extended to take account of the colon movement and of the complex illumination in the closed space of the colon.

The ultimate goal of this research is to enable the clinician to obtain histological information non-invasively, during an *in vivo* endoscopic examination. This may prevent unnecessary biopsies in a growing population of patients undergoing colon cancer screening under a recently introduced NHS programme. It may also be possible to detect early cancerous changes before they become apparent through a standard endoscopic examination.

Acknowledgements

Kevin Schomacker (MediSpectra, Inc.) is gratefully acknowledged for providing *in vivo* colon tissue spectra. We thank Dr Jon Rowe for his assistance with optimisation algorithms; and Dr Nigel Suggett and Dr Emma Hamilton (all from Birmingham University) for providing the tissue samples.

References

1. S. A. Skinner, G. M. Frydman, P. E. O'Brien, "Microvascular structure of benign and malignant tumors of the colon in humans," *Digest. Dis. Sci.* **40,** 373-84, 1995.
2. Y. Furuya ,T. Ogata, "Scanning electron microscopic study of the collagen networks of the normal mucosa, hyperplastic polyp, tubular adenoma and adenocarcinoma of the human large intestine," *Tohoku J. Exp. Med.* **169,** 1-19, 1993.
3. E. Angelopoulou, R. Molana, K. Danilidis, "Multispectral skin colour modelling," *IEEE Conference on Computer Vision and Pattern Recognition*, 2001, 635–642.
4. D. Hidovic, E. Claridge, "Modelling and validation of spectral reflectance for the colon". *Physics in Medicine and Biology* **50**, 1071-1093, 2005.
5. L. Wang S. L. Jacques, "Monte Carlo modelling of light transport in multi-layered tissues in standard c," *Univ of Texas, MD Anderson Cancer Center,* 1998.
6. Z. Ge, K. T. Schomacker, N. S. Nishioka, "Identification of colonic dysplasia and neoplasia by diffuse reflectance spectroscopy and pattern recognition techniques," *Applied. Spectroscopy* **52,** 833-9, 1998.
7. J. E. Rowe, D. Hidovic, "An Evolution Strategy using a continuous version of the Gray-code neighbourhood distribution," in *Proc. GECCO 2004, K.* Deb, ed., *LNCS* **3102,** 725-736, Springer-Verlag, 2004.
8. D. Hidovic, E. Claridge, "Model based recovery of histological parameters from multi-spectral images of the colon". *Medical Imaging 2005: Physics of Medical Imaging*, Proceedings of SPIE Vol. 5745. Flynn MJ (Ed), 127-137, 2005.
9. E. Claridge, S. J. Preece, "An inverse method for the recovery of tissue parameters from colour images" *Information. Processing in. Medical Imaging, C.* Taylor and J. A. Noble, eds., LNCS **2732,** 306-317, Springer, 2003.
10. S.J. Preece, I.B. Styles, S. Cotton, E. Claridge, A. Calcagni. Model- based parameter recovery from uncalibrated optical images. *Medical Image Computing and Computer Assisted Intervention.* LNCS vol. 3750, 509-516, 2005.

Mammographic mass eigendetection

Arnau Oliver[a,*], Jordi Freixenet[a], Robert Martí[a], Erika R.E. Denton[b], and Reyer Zwiggelaar[c]

[a]Institute of Informatics and Applications, University of Girona, 17071, Girona, Spain
[b]Department of Breast Imaging, Norfolk and Norwich University Hospital, Norwich, NR4 7UY, UK
[c]Department of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK

Abstract. A new algorithm for the detection of masses in mammographic images is presented. The algorithm has been designed in two steps. Firstly, the regions likely to be a mass are detected by using a deformable template matching approach, where the template is constructed using the eigenimages of a set of manually detected masses. Subsequently, an algorithm adapted from the eigenfaces approach is used to assure that the detected regions really depict true masses (false positive reduction). The evaluation uses a leave-one-out methodology and is based on a database of 120 mammograms, which include 40 masses and 80 normals. ROC and FROC analysis is used to demonstrate the potential of the developed approach.

1 Introduction

The proposed method is designed in two steps. Initially, an algorithm is used to detect Regions of Interest (RoI) with a high likelihood of containing a mass and, subsequently, a second algorithm is used for false positive reduction. To our knowledge, this is similar to the strategy used by radiologists: firstly they discard mammograms which certainly do not have a mass present, and in a posterior step, they might re-examine the detected suspicious regions.

Although in the literature different algorithms have been proposed for mass detection [1, 10], few of them incorporate prior knowledge of the shape of the masses. We refer to the work of Lai et al. [4], which looks for masses by using a simple circular template and a pattern matching algorithm using the normalised cross-correlation as the similarity measure. In this work we also use a template matching scheme but with two main differences. Firstly, instead of using the cross-correlation similarity, a probabilistic Bayesian scheme [3] is used. And secondly, contour information coming from an eigenanalysis of manually annotated masses is used, instead of using region information. On the other hand, some techniques have been developed, which are directly related to the false positive reduction step or RoI classification [2, 8]. In contrast with those algorithms which are task-specific, we adapted the eigenfaces approach [9], which is a well known algorithm for automatic face recognition.

2 From Eigenfaces to Eigenmasses

The original eigenfaces approach of Turk and Pentland [9] for face recognition is based on the use of the Karhunen-Loeve transform in order to find the vectors that best account for the distribution of face images (forming the face subspace) within the entire image space. The total scatter matrix (the covariance matrix) is calculated as:

$$S_t = \sum_{k=1}^{M} (x_k - \mu)(x_k - \mu)^t \tag{1}$$

where $\mu = \frac{1}{M} \sum_{k=1}^{M} x_k$ is the mean of all face samples and M the number of face images. Using the Karhunen-Loeve transform it is possible to obtain the (dimensionality reduced) eigenfaces space, as the one that maximises:

$$W_{pca} = arg \max_{W} |W^t S_t W| \tag{2}$$

There is significant parallelism between face images of the same person and mammographic RoIs images. As in face images, mammograms are rich-textured images, and it is also possible to talk about variations in illumination and pose. Changes in illumination are related to the different acquisition parameters (dose and energy of X-rays, internal density of the breast), while changes in the pose can be explained as global changes (compression related) or local changes (size and shape of the mass). To prevent these problems we firstly equalise the images using a uniform distribution model, and secondly, we cluster the RoIs of the database according to their size. Note that in contrast with face recognition, the size of the RoIs is not always the same, as it depends on the size of the (possible) mass. Although other proposals can be considered (resizing), we reported in previous work that the used solution outperforms the others [6].

To avoid confusion, note that if the RoI database only contains masses, the result of the application of such algorithm will be called *eigenmasses*, while if the database is composed of RoIs with masses and RoIs with normal tissue, we will call the resulting images *eigenrois*.

*Corresponding author, e-mail: aoliver@eia.udg.es

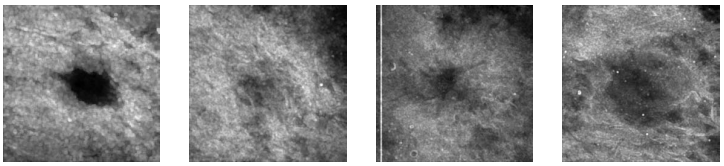


Figure 1. The probabilistic templates obtained by clustering the dataset in four clusters of various sizes. Lighter pixels represent a higher probability of a mass contour.

3 Probabilistic Mass Contour Template

The first objective of the proposal is the construction of a general mass template, which should take the shape variations into account. The main aim is that pixels with a boundary morphology which has a major representation in the database have a higher probability than pixels not belonging to it. As the template is represented as an image, the pixel brightness is associated with the probability of belonging to a contour. Thus, pixels which certainly do not represent a contour have intensity 0, pixels that always are on a contour (if any) have intensity 1, and the rest of the pixels have intermediate values.

Note that considering only the boundaries of manually segmented masses prefixes a set of contours, and contours different from them will probably be refused while, in contrast, the probability to find two masses with similar shape is very low. Thus, in order to obtain a more general template, it is constructed by looking for the sub-space that these boundaries define. This is achieved adapting the eigenfaces approach described in Section 2. Moreover, using this approach only a rough manual segmentation is needed, just including the centre and size of the mass.

With the obtained eigenmasses, it is possible to construct a probabilistic template per size. For constructing these templates, the N eigenvectors containing 95% of variation explanation were used, considering more probable shapes those with the greatest eigenvalue. Therefore, an initial template is constructed as :

$$\psi^0(x,y) = \frac{1}{N}\sum_{k=1}^{N} w_k W_k(x,y) \tag{3}$$

where $\psi^0(x,y)$ is the template, $W_k(x,y)$ is the k-th eigenmass and w_k its normalised eigenvalue (the corresponding eigenvalue divided by their sum). The contour of the eigenmasses is found by extracting the gradient from $\psi^0(x,y)$:

$$\nabla\psi^0(x,y) = \nabla\left\{\frac{1}{N}\sum_{k=1}^{N} w_k W_k(x,y)\right\} = \frac{1}{N}\sum_{k=1}^{N} w_k \nabla W_k(x,y) \tag{4}$$

This equation (image) represents the template as a weighted contours of the eigenmasses. However, in order to discard abnormal shapes, it is necessary to specify the modes of deformation of this template. Plausible shapes are those obtained from linear combinations of the eigenmass contours, and deformation will only affect the weight of the eigenvalues of each eigenmass. This is represented by a vector ξ of size N:

$$\nabla\psi^d(x,y) = \kappa\sum_{k=1}^{N}(\xi_k * w_k)\nabla W_k(x,y) \tag{5}$$

where $\psi^d(x,y)$ is the deformed template and κ is just a normalisation factor. With this definition, the vector ξ is all ones when no variations from the template occurs, and results in larger difference to the original template as it increases its values. Hence, assuming a Gaussian distribution, the probability of finding a template with such deformation is:

$$Pr(\xi) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(\xi_k - 1)\} \tag{6}$$

Note that with such definition a new parameter is included. Changes in the value of σ represent a more rigid (small σ) or a more flexible (large σ) template. Figure 1 shows the templates for four classes representing the range of mass sizes in the database.

4 Template Based Detection

Once a template per size range is constructed, the second step is to find mass boundaries (if any) in a mammogram. The developed approach is inspired by the work of Jain et al. [3], where a Bayesian inference scheme was adopted. The three following subsections describe in more detail each of the prior, likelihood and posterior terms.

4.1 Prior Distribution

The prior distribution is used to bias the global transformations (changes in translation and scale) and local deformations that can be applied to a prototype template. In contrast with the work of Jain et al. [3], rotation is not taken into account as we assume that this is represented in the probabilistic template. In this work, $\psi^{s,\xi,d}$ denotes a deformation of the original template ψ^0. This deformation is performed by locally deforming the template by a set of parameters ξ, scaling the local deformation by a factor of s, and translating the scaled version along the x and y directions by an amount $d = (dx, dy)$. Assuming that translations and scale sizes have equal probability, and using Eq. 6 for the deformation probability, the prior distribution results in:

$$Pr(s, d, \xi) = K \exp\{-\frac{1}{2\sigma^2} \sum_{k=1}^{N} (\xi_k - 1)\} \tag{7}$$

where K is a normalisation factor. Intuitively, a deformed template with a geometric shape similar to the prototype template is favoured, regardless of its size and location in the image.

4.2 Likelihood

The likelihood is a measurement of the similarity between the deformed template and the object(s) present in the image. The deformable template will be attracted and aligned to the salient edges in the input image via a directional edge potential field. For a pixel (x, y) in the input image its edge potential is defined as [3]:

$$\phi_Y(x, y) = -\exp(-\rho\sqrt{\delta_x^2 + \delta_y^2}) \tag{8}$$

where (δ_x, δ_y) is the displacement to the nearest edge point in the image, and ρ is a smoothing factor which controls the degree of smoothness of the potential field. This potential is modified by introducing a directional component relating the deformed template $\psi^{s,\xi,d}$ to the edges of the input image Y:

$$\Theta(\psi^{s,\xi,d}, Y) = \frac{1}{T} \sum_{x,y \in \psi^{s,\xi,d}} (1 + \phi_Y(x, y)|cos(\beta(x, y))|) \tag{9}$$

where the summation is over all the pixels on the deformed template, T is the number of pixels on the template, $\beta(x, y)$ is the angle between the tangent of the nearest edge and the tangent direction of the template at position (x, y), and the constant 1 is added so that the potentials are positive and take values between 0 and 1. This definition requires that the template boundary agrees with the image edges not only in position, but also in the tangent direction. Using the above energy function, the probability density of the likelihood of observing the input image, given the deformations of the template is:

$$Pr(Y|s, d, \xi) = \alpha \exp\{-\Theta(\psi^{s,\xi,d}, Y)\} \tag{10}$$

where α is a normalising constant to ensure that the above function integrates to 1. The maximum likelihood is achieved when $\Theta(\psi^{s,\xi,d}, Y) = 0$ i.e., when the deformed template $\psi^{s,\xi,d}$ exactly matches the edges in the input image Y.

4.3 Posterior Probability Density

Using Bayes rule, the posteriori probability density of the deformed template given the input image is:

$$Pr(s, d, \xi|Y) = Pr(s, d, \xi) * Pr(Y|s, d, \xi)/Pr(Y) \tag{11}$$

where $Pr(Y)$ is the normalisation factor assuring the sum of all probabilities is equal to 1. Using Eqs 7 and 10, the posterior results in:

$$Pr(s, d, \xi|Y) = K_1 \exp\{-\frac{1}{2\sigma^2}[\sum_{k=1}^{N} (\xi_k - 1) + \Theta(\psi^{s,\xi,d}, Y)]\} \tag{12}$$

As the objective is to maximise this probability, we seek to minimise the following objective function with respect to s, ξ, d:

$$\Lambda(\psi^{s,\xi,d}, Y) = \sum_{k=1}^{N} (\xi_k - 1) + \Theta(\psi^{s,\xi,d}, Y) \tag{13}$$

This function consists of two terms: a first term that measures the deviation of the deformed template from the prototype, and a second one which describes the fitness of the deformed template to the boundaries of the image.

When the algorithm has finished, a set of points of the mammogram are marked as RoIs. However, a large number of the RoIs actually correspond to normal tissue, and thus, a subsequent step is necessary in order to reduce the number of false positives. Fig. 2 shows the result of applying the algorithm to four well-defined masses.

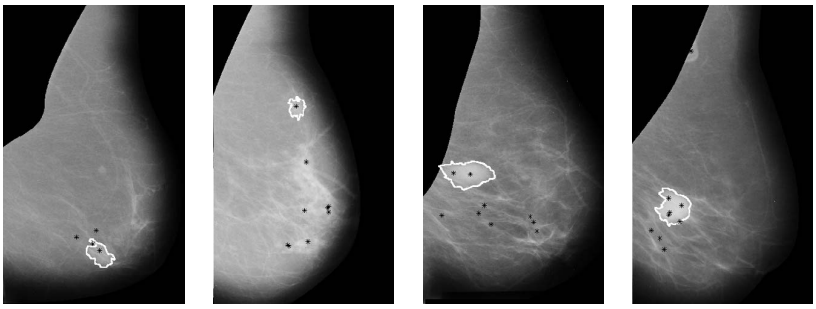


Figure 2. Suspicious regions (indicated by the star signs) found in four MIAS mammograms [7], with the annotations of the well-defined masses marked by continuous lines. Note that a posterior step is necessary to reduce the false positive regions.

5 False Positive Reduction

Once the most suspicious regions to be a mass have been found, they are cropped using the size of the template, thus obtaining a set of RoIs composed by masses and normal tissue. A classifier is needed to distinguish a RoI according to these two classes, and once again, the eigenfaces approach is used. According to this approach a database of already classified RoIs is necessary. Note that although in this work the RoI database contains only two types of RoIs, this step can be easily extended to distinguish other RoIs containing microcalcifications or architectural distortions. Thus, a model is constructed by using Eq. 2 for each RoI in the database. When a new RoI has to be tested, it will be classified as belonging to the most similar class. Although in the original algorithm the similarity was calculated using the k-Nearest Neighbour algorithm, here we used a combination of this algorithm with the ID3 decision tree. This combination results in a degree of membership for each class (see [6] for more details).

6 Results

The performance of our approach was evaluated using a total of 120 mammograms extracted from the MIAS mammographic database [7]. Of these, 40 have confirmed masses (containing most of the well-defined and spiculated masses in the database, with detailed manual outline annotations provided by an expert mammographic radiologists) while the rest were normal mammograms. The mammograms containing the well-defined masses are skewed towards the fatty background tissue MIAS classification, whilst the normal and spiculated mass containing mammograms have an even distribution over the background tissue classes. The two sets of RoIs were also extracted from these mammograms. We used four different groups for the 40 mass RoIs according to their size. Each group corresponds to the following intervals for mass sizes: $< 100\ mm^2, (100 - 180)\ mm^2, (180 - 360)\ mm^2, > 360\ mm^2$. In each interval there were, respectively, 7, 11, 12 and 7 masses, while the rest of RoIs represent normal tissue. Three masses, which were much larger than the rest, were excluded from the modelling. In addition, the original algorithm of Lai et al. [4] (as this is one of the few publications that takes shape information into account when looking for masses) has been re-implemented and included for direct comparison.

The evaluation is done using a leave-one-out methodology and Free Receiver Operating Characteristics (FROC) analysis [5]. In this work, we consider a true positive, if there is at least a 50% overlap between the manually segmented and the automatically detected regions. As is shown in Fig. 3, the proposed approaches have better performances than Lai et al.'s approach, which has a tendency to produce a large number of false positives at high sensitivity rates. Moreover, note that the inclusion of the false positive reduction step clearly reduces the number of false positives detected with the first step of the proposal.

Once the mammograms containing masses are detected, a ROC curve can be obtained measuring the accuracy with which the masses have been detected [5]. The percentage value under the curve (A_z) is an indication for the overall performance of the observer, and is typically used to analyse the performance of the algorithms. The overall performance over the 40 mammograms containing masses resulted in a A_z value of 0.84 ± 0.11 and 0.82 ± 0.13 without and with the false positive reduction step, while using the Lai et al. algorithm A_z is 0.76 ± 0.08. Thus, the proposed approaches have better performance than the original algorithm, although the false positive reduction step introduces a penalisation term in the accuracy of which the algorithm detects the masses. Table 1 shows the effect of the size of the lesion for the different algorithms. Mean and standard deviation of the values of A_z according to mass sizes are shown. Note that the Lai et al. algorithm has better performances for small masses than for larger ones. In contrast, both proposed approaches have similar performances for each class and significantly better compared to Lai et al.

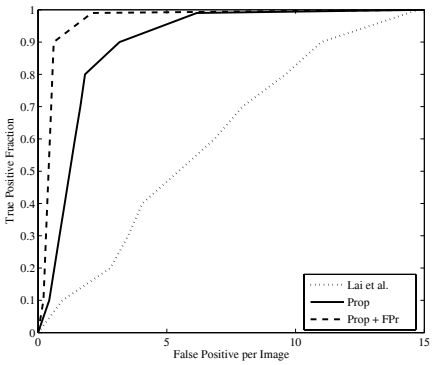


Figure 3. FROC analysis of the algorithm over the set of 120 mammograms. It is clear that both proposals outperform the original one.

	< 100	$100 - 180$	$180 - 360$	> 360
A_z **(Lai)**	0.80 ± 0.05	0.78 ± 0.08	0.74 ± 0.09	0.71 ± 0.05
A_z **(Eigen)**	0.85 ± 0.13	0.83 ± 0.11	0.82 ± 0.09	0.87 ± 0.06
A_z **(Eigen & FPr)**	0.83 ± 0.09	0.81 ± 0.14	0.80 ± 0.16	0.80 ± 0.07

Table 1. Influence of the lesion size (in mm^2) for the algorithm of Lai et al. and the proposed methods (second row without and third row with false positive reduction (FPr)). The results show mean and standard deviation A_z values.

7 Conclusions

The aim of this work is the development of an algorithm capable of finding masses using shape information. The approach is divided into two main algorithms. Firstly, using a database of RoIs containing masses, a probabilistic template is created representing the most probable contours of a mass. This template forms the basis to search for masses in a mammogram using a probabilistic scheme. The result is a set of RoIs containing suspicious regions. Secondly, a false positive classification algorithm has been developed to reduce the number of RoIs not depicting masses. The performance of both approaches was evaluated using a leave-one-out methodology and FROC and ROC analysis. Both stages outperform the work of Lai et al. These results are promising and further work will be focused on: a) a feature selection process to further reduce the number of false positives without penalising the accuracy of the Bayesian template matching, b) the inclusion of information related to the parenchymal tissue distribution (i.e. grey-level and texture information), and c) the use of larger databases to test the clinical relevance of the work.

Acknowledgments

This work was partially supported by MEC grant nb. TIN2005-08792-C03-01.

References

1. D. Brzakovic, X.M. Luo, and P. Brzakovic. An approach to automated detection of tumors in mammograms. *IEEE Trans. Med. Imag.*, 9(3):233–241, 1990.
2. Y.H. Chang, L.A. Hardesty, C.M. Hakim, T.S. Chang, B. Zheng, W.F. Good, and D. Gur. Knowledge-based computer-aided detection of masses on digitized mammograms: A preliminary assessment. *Med. Phys*, 28(4):455–461, 2001.
3. A.K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(3):267–278, 1996.
4. S.-M. Lai, X. Li, and W.F. Bischof. On techniques for detecting circumscribed masses in mammograms. *IEEE Trans. Med. Imag.*, 8(4):377–386, 1989.
5. C.E. Metz. Evaluation of digital mammography by ROC analysis. In *Int. Work. on Dig. Mammography*, pages 61–68, 1996.
6. A. Oliver, R. Martí, J. Martí, A. Bosch, and J. Freixenet. A new approach to the classification of mammographic masses and normal breast tissue. In *IAPR Int. Conf. on Patt. Rec.*, page to appear, Hong Kong, August 2006.
7. J. Suckling, J. Parker, et al. The Mammographic Image Analysis Society digital mammogram database. In *Int. Work. on Dig. Mammography*, pages 211–221, 1994.
8. G.T. Tourassi, R. Vargas-Vorecek, D.M. Catarious, and C.E. Floyd. Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information. *Med. Phys*, 30(8):2123–2130, 2003.
9. M.A. Turk and A.P. Pentland. Eigenfaces for recognition. *J. Cogn. Neuroscience*, 3(1), 1991.
10. R. Zwiggelaar, T.C. Parr, J.E. Schumm, I.W. Hutt, C.J. Taylor, S.M. Astley, and C.R.M. Boggis. Model-based detection of spiculated lesions in mammograms. *Med. Image Anal.*, 3(1):39–62, 1999.

Segmentation of Brain MRI in Young Children

Maria Murgasova[a], Leigh Dyet[b], David Edwards[b], Mary Rutherford[b],
Jo Hajnal[b] and Daniel Rueckert[a]

[a]Visual Information Processing Group, Department of Computing, Imperial College London
[b]Department of Imaging Sciences, Faculty of Medicine, Imperial College London

Abstract. This paper describes an automatic tissue segmentation algorithm for brain MRI of young children. Existing segmentation methods developed for the adult brain do not take into account the specific tissue properties present in the brain MRI of the young children. We examine the suitability of the state-of-the-art methods developed for the adult brain when applied to the segmentation of the young child brain MRI. We describe a method of creation of a population-specific atlas from young children using a single manual segmentation. The method is based on non-linear propagation of the segmentation into population and subsequent affine alignment into a reference space and averaging. By using this atlas we improve the performance of the popular EM segmentation algorithm on brain MRI of young children.

1 Introduction

The problem of automatic segmentation of brain MRI has been extensively studied in past decade thanks to its applications in clinical studies of normal and diseased brain. Recent research of brain development in prematurely born children requires reliable and accurate automatic segmentation techniques for the neonatal and early childhood brain as well. Approximately 40 % of prematurely born children have cognitive, neurological or behavioural impairment [1] and the cerebral abnormalities underlying problems such as minor motor impairment or inattention still remain unclear. Accurate tissue quantification of the preterm brain in early childhood would allow us to study the growth of different structures and to try to relate developmental outcomes to changes in volume and shape of anatomical structures.

Current state-of-the-art methods for automatic segmentation of the adult brain into 3 basic tissue classes (white matter, grey matter and cerebrospinal fluid) typically rely on single voxel intensity classification spatially constrained by a probabilistic atlas ([2], [3], [4], [5], [6], [7]). This class of methods is sensitive to overlap in the tissue intensity distributions and therefore the probabilistic atlas is often used not only to initialize, but also to constrain the segmentation process, usually in form of a-priori information ([2], [3], [5], [6]). In addition to noise, intensity non-uniformity and partial volume effect which cause such overlap in the adult brain, neonatal and infant brain tissue intensities have much greater natural variation caused by on-going process of myelination of white matter. This is most profound in neonates where white matter is extremely difficult to distinguish from grey matter on T1 weighted MRI sequences. By the age of 1 or 2 years, the majority of white matter has myelinated. However, the natural tissue intensity distribution overlap results in large misclassifications in central brain structures. This problem is further compounded by using an adult atlas as a spatial prior as the anatomy of the brain in early childhood differs significantly from adult in central structures of the brain [8].

In this paper we focus on the segmentation of brain MRI of 1 and 2-year-old children. We analyze tissue properties of the brain in early childhood and examine the suitability of state-of-the-art methods developed for segmentation of the adult brain when applied to young children. We show how to bootstrap a probabilistic tissue atlas for a specific population based on registration of a single manual segmentation to a population and subsequent alignment to the space of a reference subject and averaging. We demonstrate that the results of the expectation-maximization method by Van Leemput *et al.* ([2], [3]) on young children can be significantly improved by using an atlas which appropriately captures the anatomical variability of the infant brain and compare the results with a simple version of a method inspired by Tasdizen *et al.* [9].

2 Related Work

To date the research in MRI brain segmentation has focused on adult and neonatal brains. The most popular class of tissue classification methods is based on expectation-maximization framework ([2], [3], [4], [5]). These methods usually use a probabilistic atlas to initialize the segmentation and then iteratively interleave calculation of intensity probability distribution for each tissue class and/or bias field, and refinement of spatial probability maps. A mixture Gaussian distribution is assumed as a model for the voxel intensity probability distribution as it is easy to incorporate into EM framework and reasonably describes the real tissue intensity distributions. However, this assumption results

in some misclassification as the real distribution is not strictly Gaussian. This is most pronounced in CSF distribution as very large boundary with GM results in distribution distortion thanks to partial volume effect, see Fig. 1. Wells *et al.* [5] suggest to estimate the real distributions instead of the means and variances of the Gaussian distributions. This extended method is not an instance of EM algorithm but it is still reported to be robust in practice. Some recent studies propose to improve the segmentation results by including a model for the partial volume effect in the EM framework [10].

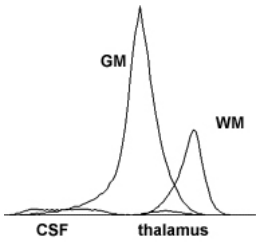


Figure 1. Histogram of the tissue intensity probability distribution for white matter, grey matter, cerebrospinal fluid and thalamus (a deep grey matter structure) based on a manual segmentation of brain MRI from a 2-year-old child. Distributions of the grey matter and csf differ significantly from the Gaussian distribution thanks to partial volume effect. The intensity distribution of thalamus lies in the region of the overlap of the white and grey matter distributions.

Inevitable misclassifications resulting from overlaps in tissue intensity distributions can be partially avoided by spatially constraining the segmentation process with prior information in form of an aligned probabilistic atlas at each iteration ([2], [3]); however this makes the method very sensitive to the correct alignment of the atlas with the image. This is almost impossible for subjects with distorted anatomy or child brains which significantly differ from adult brains [8]. To solve this problem, [11] and [12] suggest to warp the atlas to subject by non-rigid registration. In [13] Pohl reports difficulties with non-rigidly registering the template to image which is consistent with our own experience. The problem of correct alignment of the atlas with image can be overcome by iteratively warping a deformable atlas (a subject with attached labeling) and refining the segmentation at the same time [14], [6].

An alternative non-parametric method for adult brain classification was developed by Cocosco *et al.* [7]. In this method the atlas is used only to select the training samples which are consequently used in kNN classification. This method is reported to deal well with anatomies which differ from the prior information. However, it cannot deal with the problem of tissue intensity distribution overlap in single-channel T1 data. Similarly, Tasdizen *et al.* [9] use the atlas only to initialize the estimation of multi-dimensional distribution of 7-neighbourhood. Their method relies on observation that these 7-voxel samples are concentrated in clusters.

As mentioned above the adult brain segmentation methods were not developed to deal with the challenges of segmenting the developing brain (except from the semi-automatic method by Warfield *et al.* [6] which has been applied to neonatal brain images). However, the principles still can be used when tuned to specific properties of the neonatal and infant brain. This was shown by Prastawa *et al.* [15] who used the EM approach ([2], [3]) with an atlas derived from single subject segmentation used as a prior, followed by non-parametric classification similar to [7] on neonatal data.

3 Methods

3.1 Expectation-Maximization Algorithm

For segmenting the T1-weighted images from 1- and 2- year old children we have chosen the EM algorithm by Van Leemput *et al.*([2], [3]). We aligned the CCHMC pediatric brain atlas created from population of 5 to 9-year-old children [8] (Fig. 4c) with the images. All data acquired on 1.0T HP MRI scanner was corrected by the N3 software [16] in preprocessing step which proved to be sufficient as the bias field was not very strong. For the images acquired on 0.5T Apollo MRI scanner no bias field correction was necessary. This allowed us to study the EM method without the bias correction step and consequently better analyze its effect on the resulting segmentation.

The EM algorithm is a general technique for finding missing data based on observed data and maximum likelihood parameters estimates. In our case the observed data are the image intensities, the missing data are the labels and the parameters are the means and variances of the Gaussian distribution which is assumed for the intensity distribution of each tissue class. This is an iterative process which interleaves the calculation of posterior probabilities of each voxel belonging to each tissue class (white matter, grey matter, cerebrospinal fluid, other) - the expectation step, with maximum likelihood estimation of the Gaussian distribution parameters - the maximization step.

E-step:

$$p_{ik}^m = \frac{G(y_i, \mu_k^{m-1}, \sigma_k^{m-1}) p_{ik}^{atlas}}{\sum_j G(y_i, \mu_j^{m-1}, \sigma_j^{m-1}) p_{ij}^{atlas}}$$

M-step:

$$\mu_k^m = \frac{\sum_i p_{ik}^m y_i}{\sum_i p_{ik}^m} \quad \sigma_k^m = \frac{\sum_i p_{ik}^m (y_i - \mu_k^m)^2}{\sum_i p_{ik}^m}$$

where p_{ik}^m denotes the posterior probability of i^{th} voxel belonging to tissue k at m^{th} iteration, y_i is an intensity of i^{th} voxel and $G(.,.,.)$ denotes the Gaussian distribution with the mean μ_k^m and variance σ_k^m.

We found that this method significantly overestimates white matter in the central structures of the brain. This is most profound in the thalamus, a deep grey matter structure, where about 30% of voxels were misclassified as white matter (see Fig. 2 and Fig. 4b). The histogram based on the manual segmentation of a 2-year-old 1.0T data set (Fig. 1) shows that the voxel intensities of this structure lie in the region of the overlap of white and grey matter intensity distributions and therefore the classification depends on the prior information. However, the prior information from the probabilistic atlas [8] overestimates the white matter in this area as it does not describe the anatomy of the brain in early childhood adequately (Fig. 4c).

(a) (b) (c) (d)

Figure 2. Misclassification in thalamus: Substantial amount of deep grey matter classified as white. (a) MR image of a 2-year-old brain (b) EM segmentation using the adult atlas (c) EM segmentation using the 2-year-old child atlas (d) manual segmentation

3.2 Bootstrapping the Prior Information for 1 and 2-year-old Brain

We obtained a manual segmentation of a single 2-year-old subject to create a probabilistic atlas of the 1 and 2-year-old brain. The manual segmentation was then transformed to 37 2-year-old subjects by performing non-rigid registration [17]. All 37 segmented subjects were then aligned with the reference subject by affine registration and the segmentations were averaged. The resulting probability maps describe the anatomy of the 2-year-old brain correctly when aligned with a subject by affine registration.

We also created a separate probabilistic atlas of the 1-year-old brain. First we non-rigidly warped the manual segmentation of a 2-year-old brain to the scan of the same subject at 1 year to minimize the registration error. The same process of the atlas creation was then repeated with 34 1-year-old scans.

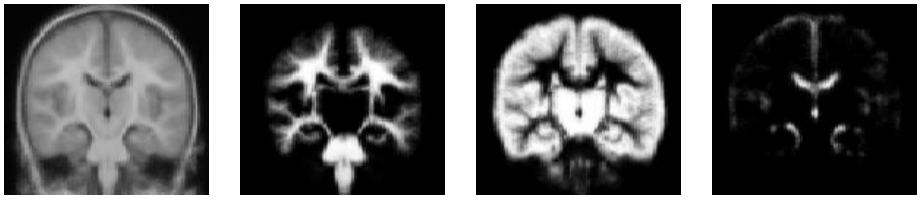
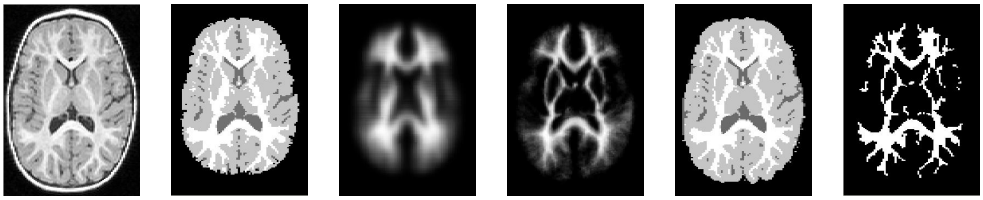


Figure 3. The probabilistic atlas created from population of 2-year-old children.

We were also able to use the single subject segmentation as the prior information by registering it non-rigidly to the image and subsequent blurring to accommodate anatomical variability.

Both of these methods resulted in significant improvement in segmentation of the central brain structures (Fig. 2 and 4). However, the white matter was still consistently overestimated which is caused by trying to separate the tissue classes based on single voxel intensity on global scale.



(a) (b) (c) (d) (e) (f)

Figure 4. Improvement of segmentation using our new atlas: (a) MR image of a 2-year-old brain (b) segmentation using the 5 to 9-year-old CCHMC brain atlas (c) the 5 to 9-year-old CCHMC brain atlas (d) our 2-year-old brain atlas (e) segmentation using our 2-year-old brain atlas (f) manual segmentation

3.3 Local Distribution Estimation Method

Young child brains exhibit larger natural intensity variability within the tissue classes. However, the boundaries are often visible to human observer. Therefore local estimation of tissue distributions can improve the segmentation. Inspired by [9] we implemented a local distribution estimation method which we have compared with the results of the EM approach in previous sections. Non-linearly warped manual segmentation was used as prior information and the tissue intensity distributions were then estimated by Parzen windowing for each voxel separately after masking with local Gaussian kernel. Each voxel was then assigned to the tissue class with the highest probability. Even though this approach improved the segmentation in small detail it was also less stable and sensitive to noise than the methods which estimate the tissue intensity distributions globally.

4 Validation and Results

4.1 Validation Using Manual Segmentation

First we validated our results on a single subject for which we obtained the manual segmentation. Validation was performed only in the brain area, excluding skull, brain stem and cerebellum. We segmented the image with our implementation of the EM method (section 3.1) with 3 different atlases: our new probabilistic atlas based on population of 2-year-old children (2yr atlas), blurred manual segmentation (2yr seg) and the standard adult atlas obtained from SPM software [18], [19] (adult atlas). For comparison we also segmented the image with EMS software [20] developed by Van Leemput. This software uses the SPM adult atlas, and unlike in our implementation, the bias correction is incorporated within the EM framework. We validated the approach with and without Markov Random Fields (MRF). Segmentation by local method was evaluated as well. The manual segmentation was used as the ground truth. The correctness of the segmentation was calculated using Dice metric

$$D = \frac{2 * |T_{gt} \cap T_{seg}|}{|T_{gt}| + |T_{seg}|}$$

where T_{gt} and T_{seg} denote the set of samples of given tissue type in the ground truth and the automatic segmentation, respectively. The validation results (Fig. 1a) show that our new atlas improved the segmentation of both white and grey matter by 4% compared to EM segmentation with the adult atlas by both EMS and our implementation. The improvement is most significant in the thalamus where performance increased by 12% in comparison with EMS and by 26% compared to our implementation. Better performance of the EMS software in comparison with our implementation of EM with the adult atlas can be explained by the fact that the bias correction step partially compensates for natural intensity variation in the infant brain. When a blurred manual segmentation was used as prior information for EM, the performance slightly increased, but in this case it did not have to be non-linearly warped to the subject and therefore the results do not include the registration error. Overall performance of the local method was only marginally better than EMS thanks to sensitivity to noise and instability. However, it resulted in 94% correct segmentation of the thalamus. Again, the results do not include the registration error.

4.2 Two step validation

In the second experiment we compared our 3 methods on 37 2-year-old subjects. As we do not have the ground truth for any other subject we performed a two step segmentation. First, we segmented all the subjects by each of our methods and the resulting probability maps were used as prior information for segmentation of the manually segmented subject.

	WM	GM	CSF	thalamus
EMS	0.844	0.883	0.694	0.779
EMS-MRF	0.855	0.891	0.616	0.799
EM adult atlas	0.847	0.879	0.672	0.664
EM 2yr atlas	0.885	0.922	0.784	0.923
EM 2yr seg	0.904	0.927	0.814	0.926
local	0.877	0.906	0.625	0.939

(a)

	WM	GM	CSF	thalamus
EM 2yr atlas	0.716	0.852	0.569	0.878
EM 2yr seg	0.872	0.907	0.740	0.847
local	0.855	0.809	0.404	0.916

(b)

Table 1. (a) Performance of the segmentation methods on a single 2-year-old subject (b) Two-step validation of the segmentation methods on 37 2-year-old subjects

The segmentations were then compared with the manual segmentation using Dice metric (Fig. 1b). In this experiment the EM segmentation with non-rigidly warped blurred manual segmentation performed best overall whereas the local method was the most successful in the thalamus area.

5 Conclusion

The more challenging problem of segmentation of the brain in early childhood helped us to understand the strengths and weaknesses of different approaches to brain MRI tissue classification. We showed how to create a population specific atlas and consequently improve the performance of the popular EM approach. However, more research needs to be done in this area to produce satisfactory results for clinical studies. In future we would like to focus on development of a robust method which would perform tissue classification locally and incorporate iterative non-rigid warping of deformable atlas into segmentation process.

References

1. N. Marlow, D. Wolke, M. A. Bracewell et al. "Neurologic and Developmental Disability at Six Years of Age after Extremely Preterm Birth." *Pediatrics* **115(2)**, pp. 286–94, February 2005.
2. K. V. Leemput, F. Maes, D. Vandermeulen et al. "Automated Model-Based Tissue Classification of MR Images of the Brain." *IEEE Trans. Med. Imaging* **18(10)**, pp. 897–908, 1999.
3. K. V. Leemput, F. Maes, D. Vandermeulen et al. "Automated model-based bias field correction of MR images of the brain." *IEEE Trans. Med. Imaging* **18(10)**, pp. 885–896, 1999.
4. J. Ashburner. *Computational Neuroanatomy*. Ph.D. thesis, University College London, 2000.
5. W. M. Wells III, W. E. L. Grimson, R. Kikinis et al. "Adaptive segmentation of MRI data." volume 15, pp. 427–441. 1996.
6. S. K. Warfield, M. Kaus, F. A. Jolesz et al. "Adaptive template moderated spatially varying statistical classification." In *MICCAI '98: Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 431–438. Springer-Verlag, London, UK, 1998.
7. A. P. Z. Chris A. Cocosco & A. C. Evans. "A fully automatic and robust brain MRI tissue classification method." In *Medical Image Analysis*, volume 7, pp. 513–527. 2003.
8. S. H. M. Wilke, V.J. Schmithorst. "Normative pediatric brain data for spatial normalization and segmentation differs from standard adult data." *Magnetic Resonance in Medicine* **50(4)**, pp. 749–757, 2003.
9. T. Tasdizen, S. P. Awate, R. T. Whitaker et al. "MRI Tissue Classification with Neighborhood Statistics: A Nonparametric, Entropy-Minimizing Approach." In *MICCAI (2)*, pp. 517–525. 2005.
10. N. Joshi & M. Brady .
11. K. Pohl, W. Wells, A. Guimond et al. "Incorporating non-rigid registration into expectation maximization algorithm to segment MR images." In *Medical Image Computing and Computer-Assisted Intervention*, volume 2488 of *Lecture Notes in Computer Science*, pp. 564–572. Springer-Verlag, Tokyo, Japan, 2002.
12. E. D'Agostino, F. Maes, D. Vandermeulen et al. "Non-rigid atlas-to-image registration by minimization of class-conditional image entropy." In *MICCAI (1)*, pp. 745–753. 2004.
13. K. Pohl. *Prior Information for Brain Parcellation*. Ph.D. thesis, Massachusetts Institute of Technology, 2005.
14. J. Ashburner & K. Friston. "Unified segmentation." *NeuroImage* **26**, pp. 839–851, 2005.
15. M. Prastawa, J. H. Gilmore, W. Lin et al. "Automatic Segmentation of Neonatal Brain MRI." In *MICCAI (1)*, pp. 10–17. 2004.
16. A. E. A. Sled, J.G.; Zijdenbos. "A nonparametric method for automatic correction of intensity nonuniformity in MRI data." *Medical Imaging, IEEE Transactions on* **17(1)**, pp. 87–97, 1998.
17. D. Rueckert, L. I. Sonoda, C. Hayes et al. "Non-rigid registration using free-form deformations: Application to breast MR images." *IEEE Transactions on Medical Imaging* **18(8)**, pp. 712–721, 1999.
18. Statistical Parametric Mapping. "http://www.fil.ion.ucl.ac.uk/spm/."
19. R. Frackowiak, K. Friston, C. Frith et al. *Human Brain Function*. Academic Press, second edition, 2003.
20. Expectation Maximization Segmentation. "http://www.medicalimagecomputing.com/EMS/."

Wireless Capsule Endoscopy video segmentation using Support Vector Classifiers and Hidden Markov Models

Michal Mackiewicz[a]*, Jeff Berens[a], Mark Fisher[a]

[a]School of Computing Sciences , University of East Anglia, Norwich, UK,

Abstract. Wireless Capsule Endoscopy is a colour imaging technology that enables close examination of the interior of the entire small intestine. In this paper we present a method of automatically discriminating oesophagus, stomach, intestine and colon tissue which can significantly speed-up video analysis time. Our approach first divides the WCE image into sub-images and rejects all sub-images in which tissue is not clearly visible. We then create a feature vector combining colour and texture information of the entire image and valid sub-images. Colour features are derived from Hue Saturation histograms, compressed using a hybrid transform, incorporating the Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA). A second feature combining colour and texture information is derived using Local Binary Patterns. The video is segmented into meaningful parts using Support Vector Classifier built within the framework of a Hidden Markov Model(HMM). We present experimental results that demonstrate the effectiveness of this method.

1 Introduction

Examination of the entire of the small intestine was until very recently a difficult procedure. The solution to this problem first appeared in [1], and involves the use of wireless transmission to send images from inside the small intestine to the outside world. The capsule (manufactured by Given Imaging Ltd. [2]) is swallowed and propelled through the food tract by normal peristalsis taking 2 pictures ($256 * 256$ pixels) a second. These images are relayed via a transmitter to a recorder worn by the patient on a belt. At the end of the 8 hour battery lifetime, the image data can be uploaded to a workstation for later viewing. The stored data consist of $\sim 50,000$ images, and is viewed as a video sequence using software provided by the manufacturers. An important factor with regard to using the WCE system is that viewing each video requires a significant time-commitment and the close concentration of an expert clinician. Even for an experienced viewer it can take over an hour to analyse a WCE video [3]. Considering the hourly wages of specialists who view WCE videos, one can easily guess that capsule investigation is a very expensive procedure.

There has been little research as far as WCE image analysis is concerned. The manufacturers provide only one image analysis function to the their software: The Suspected Blood Indicator, designed to find the location of bleeding areas. There is also a localisation function which displays the relative position of the capsule on a graphical torso model. In order for these functions to work, the viewer must find and mark two important locations in the video, firstly, at the Pylorus and secondly, at the Ileocaecal Valve (IV). These points mark the extent of the small intestine, and must be input manually. Finding the place in the video where the WCE leaves the stomach can be difficult and time-consuming, even for an experienced viewer and annotating the place where the capsule passes the IV is even more difficult since intestine and colon tissues are very similar and are often contaminated with faecal material that occludes the camera view. Providing those locations automatically would significantly reduce the amount of time taken by a doctor, when assessing the WCE video. In [4] we proposed stomach/intestine/colon discriminators. In [5] we presented one more discriminator namely mouth/oesophagus&stomach and suggested using the appropriate discriminators to estimate Gastric Transit Time and Intestinal Transit Time, which are useful cues for clinicians. Our discriminators utilize hue saturation chromaticity histograms which are compressed using a *hybrid transform* [6], incorporating the Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA).

Coimbra et al. [7] present an attempt to segment a WCE video into meaningful parts. They divide the video into four zones: Entrance ($Z1$) - consisting of image frames acquired from the mouth and oesophagus as well as those acquired before the capsule is swallowed; Stomach ($Z2$) - whose limits are determined by the oesogastric junction and the pylorus; small intestine ($Z3$) - delimited by the pylorus and IV; Colon ($Z4$) - from IV to the end of the footage. MPEG-7 descriptors (Scalable Colour and Homogeneous Texture) are used as low-level image features. The classification is performed using a Bayesian classifier, which assigns a topographic location label to each frame in the video. Iteration is used to minimise the segmentation error, resulting in three parameters that show the positions of transitions between the four previously defined zones. This method is similar to work we have presented in [4, 5]. The tissue classifier described above correspond to the tissue discriminators described in our work. Care should be taken when comparing these results: although our results are superior to those presented in [7], they were obtained using different video sequences.

*mm@cmp.uea.ac.uk

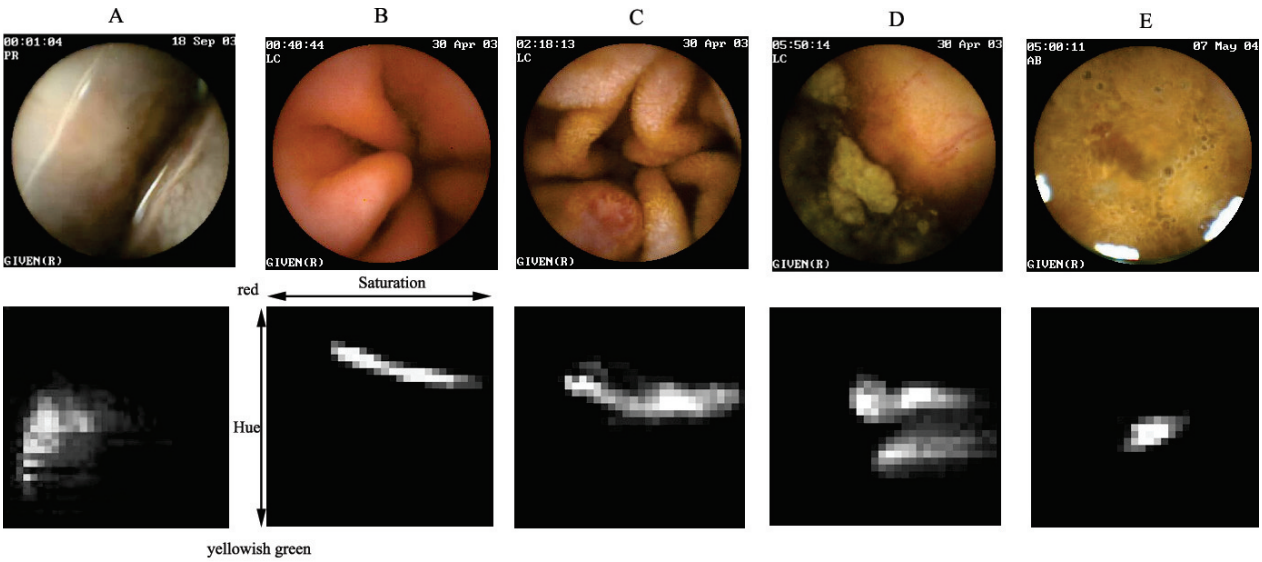


Figure 1. A figure showing WCE images taken from A) mouth B) stomach, C) small intestine, D) partially faeces occluded colon and E) completely occluded colon; and below their corresponding equalized HS histograms.

The rest of the paper is organised as follows. Section 2 describes the image features that we use for video segmentation. In Section 3, the construction of the single frame Support Vector classifier is explained. The sequence of classification results constitutes the input for the Hidden Markov Model which is described in Section 4. Section 5 discusses the experiments and Section 6 draws the conclusions.

2 Feature extraction

The distribution of colours in an image often provides a useful cue for image indexing and object recognition. The most commonly used method of representing image colour information is the colour distribution histogram, which is relatively invariant to image scale changes, translation and rotation about the viewing axis, and partial occlusion [8]. WCE images might be classified using this method since colour information is the primary feature analysed by the clinician. We convert the WCE RGB images into the HSI colour space and, ignoring the intensity information, we form 2-D HS histograms. There is a great deal of intensity variation in WCE images as the distance between the WCE and the intestine surface constantly varies. By ignoring intensity information we force intensity invariance, and also reduce data size. The range of colours present in WCE images is relatively small, mapping to a region covering just around 20% of the possible colour space, and so we equalise the histograms to this subset (of red to yellowish-green) colours. Figure 1 shows typical WCE images and their histograms. It can be seen that the colour distribution of the stomach is slightly shifted towards red compared to the intestine distribution. It is also clear that the colon tissue colour distribution is highly similar to that of the small intestine. However, it is possible to identify colon tissue by the presence of faeces - typically most of the colon is completely obscured by faecal contamination with its distinguishing hue-saturation signature. The 2-D histogram feature vector is a large structure, and for practical applications compression is necessary. Consequently, we apply a hybrid transform consisting of DCT compression followed by PCA, which was shown to perform better than DCT or PCA alone [6]. The hybrid transform allows a small feature vector that provides a fast and accurate model of the histogram data.

Texture has been often used as a feature in medical image analysis techniques. In most medical imaging systems, textural features are calculated from a grey-scale image. WCE is a colour imaging technology and therefore, more recent methods involving colour image texture can be employed. The Local Binary Pattern operator (LBP) [9, 10], is one of many colour texture methods that has proven to perform well in texture classification applications. In this work, we use a method of calculating LBP recently introduced by Connah and Finlayson [11], which characterises the images by their 3D LBP histograms. This method is an extension of the work of Mäenpää and Pietikäinen [9], who computed LBP histograms in three colour channels independently. The results will reflect that computing a 3D histogram retains more information about the image, which yields superior performance. Similarly to HS histograms, we apply Principal Component Analysis, which reduces the feature vector length and in some cases improves the results.

WCE images are often obscured (to a varying degree) by strong shadows, or by air-bubbles and other artifacts - such

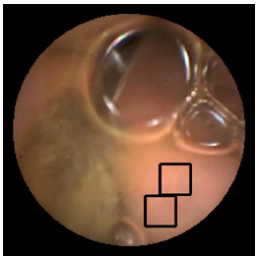


Figure 2. SubIR in a sample WCE image.

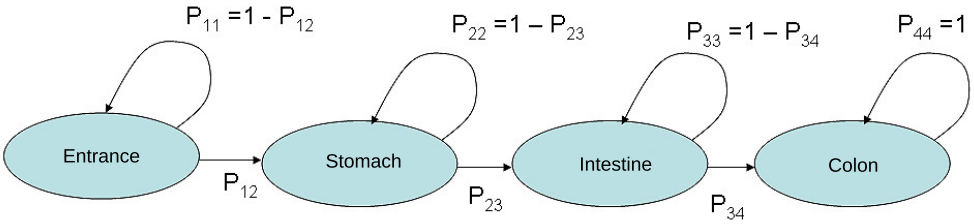


Figure 3. A figure showing the HMM for capsule video segmentation

as mucus, bile, faeces, food etc. Histograms built using the entire image will contain any visual contamination present in the image. To address this problem, we want to extract only those parts of the image which contain only visible tissue. In order to do this, we divide each WCE image into a grid of 28 sub-images, 32*32 pixels each, covering most of the image area. We derive five parameters for each of the sub-images - Mean Intensity, Saturation, Hue, and Standard Deviation of Intensity and Hue. The values for these parameters were set by experiment so that sub-images containing visual contamination (i.e. outside the expected colour range for the tissue type) are rejected. We test each sub-image against the five parameters and discard those that fall outside the range of values typical for visually clear images of gastrointestinal tissue. The remaining sub-images form a sub-image region (SubIR) that will be used in a feature extraction process. Figure 2 shows a typical image acquired in the stomach only those sub-images selected by the algorithm described above. Consequently, the features are extracted from the entire image as well as from SubIR segmented in the manner described above.

3 Support Vector Classifier

The features are classified into four classes (entrance, stomach, intestine and colon) using Support Vector Classifier (SVC) [12]. The disadvantage of SVCs is the fact that they do not provide posterior class probabilities which are needed in our application in the video segmentation stage as the input of the Hidden Markov Model. In the literature we can find a few approaches aiming to approximate these probabilities [13, 14]. Platt in [13] proposes to approximate them by a sigmoid function. Our application uses the algorithm proposed by Lin et al. [14], which is the improved version of the Platt's original algorithm. The number of classes in our problem is four, thus the use of multi-class classifier is necessary. Our approach uses the widely used multi-class classifier design called *pairwise coupling* [15] which combines together all pairwise comparisons for each pair of classes obtained using binary SVC classifiers.

4 Hidden Markov Model (HMM) Video Segmentation

HMMs [16] are widely used in video segmentation. In [17], HMM is used to segment the video into separate shots. The transitions between shots are modelled using different states: fade, dissolve, cut, pan or zoom. Figure 3 shows the HMM used for capsule endoscopy video segmentation. We use a simple left-to-right model [16] consisting of four states: entrance, stomach, small intestine and colon. Each arc in Figure 3 has an associated probability of either a transition to the next state or remaining in the same one. The input of the HMM is the sequence of conditional probabilities produced by the SVC. The model was built from every tenth frame of each video. The state transition matrix is learned during the training phase. Having trained the state transition matrix, the standard Viterbi algorithm is used to estimate the most probable sequence of states. Those states are time-aligned with the feature sequence that generated them. Once the transitions between states have been found, the video can be segmented into respective parts.

5 Experiments and Results

We were given 76 annotated capsule endoscopy videos from the Norfolk & Norwich University Hospital. The videos were annotated by the clinician into meaningful parts: entrance, stomach, intestine and colon. In our first experiment, we tested three single image SVC classifiers namely: entrance/stomach, stomach/intestine and intestine/colon. From each video we randomly selected 1000 images from each region. Next, we trained each classifier using 10000(2*5000) frames randomly selected from the above image set. Three kernels

Table 1. Feature vectors tested in the first stage

HS histogram (HS)
LBP histogram (LBP_3D)
1D LBP histogram (LBP_1D)
3D LBP and HS histograms (LBP_3D_HS)

(linear, radial and polynomial) were tested and ten-fold cross-validation was performed. The HS histograms were quantized into $32 * 32 = 1024$ bins. We applied the DCT to the histograms and carried out the PCA on the first 136 DCT coefficients. LBP histograms were built using 8 sampling points which gives us 7 unique patterns, $21(3 * 7)$ bins for the independent 1D histogram and $343(7^3)$ for the joint 3D histogram.

The first experiment consisted of two stages. In the first stage we tested only feature vectors extracted from the entire images (see Table 1), in the second we narrowed our experiment only to frames where at least one SubIR can be found (see Table 3). Consequently, in the second stage the features were extracted from the entire images (as in the first stage) as well as from SubIRs. The aim of this experiment was to find the set of features which yield the best classification results, hence we tested a number of feature vector combinations. In all cases histograms were compressed using PCA (LBP) and *hybrid transform* (HS). The results (see Table 2) clearly indicate that combining features improves the classification accuracy. In both stages of the experiment the best results were reached by the classifiers combining all the available features.

Table 2. Percentage of correct classifications

Classifier	entrance/stomach	stomach/intestine	intestine/colon
First stage classification results			
HS	94.5	92.8	84.5
LBP_3D	97.4	93.2	81.9
LBP_1D	93.5	84.7	81.8
LBP_3D_HS	**98.3**	**95.9**	**87.2**
Second stage classification results			
LBP_ent	97.3	93.8	78.6
LBP_reg	94.4	88.7	77.3
HS_ent	89.6	92.0	80.5
HS_reg	87.5	90.7	71.1
LBP_ent_reg	97.8	94.1	81.0
HS_ent_reg	94.5	93.0	82.1
LBP_ent_reg_HS_ent	98.8	96.2	87.3
LBP_reg_HS_ent_reg	97.4	94.5	85.9
LBP_ent_reg_HS_reg	99.0	95.5	83.4
LBP_ent_HS_ent_reg	99.0	**96.4**	87.5
LBP_ent_reg_HS_ent_reg	**99.4**	96.0	**87.9**

Table 3. Feature vectors tested in the second stage

HS histogram formed from the entire image (HS_ent)
HS histogram formed from the SubIR (HS_reg)
3D LBP histogram formed from the entire image (LBP_ent)
3D LBP histogram formed from the SubIR (LBP_reg)
3D LBP histograms formed from 1) the entire image and 2) SubIR (LBP_ent_reg)
HS histograms formed from 1) the entire image and 2) SubIR (HS_ent_reg)
3D LBP and HS histograms formed from the entire image and 3D LBP histogram formed from the SubIR (LBP_ent_reg_HS_ent)
3D LBP and HS histograms formed from the SubIR and HS histogram formed from the entire image (LBP_reg_HS_ent_reg)
3D LBP and HS histograms formed from the SubIR and 3D LBP histogram formed from the entire image (LBP_ent_reg_HS_reg)
3D LBP and HS histograms formed from the entire image and HS histogram formed from the SubIR (LBP_ent_HS_ent_reg)
3D LBP and HS histograms formed from 1) the entire image and 2) SubIR (LBP_ent_reg_HS_ent_reg)

In the second experiment, we tested the performance of HMM discriminators. Similarly to the first experiment, we trained the set of SVC classifiers (see Table 4). Next, we classified each tenth frame of the video using these classifiers, calculated the conditional probabilities and computed the best possible sequence of states using Viterbi algorithm. Discriminators 1,3 and 5 were built only from one classifier which classifies each frame from the classification sequence, whereas discriminators 2,4 and 6 use two classifiers, one classifies only the frames containing SubIRs and the other the rest of the frames. The accuracy of the algorithm was assessed by the frame difference (error) between the point in the video where the boundary has been manually annotated (by a clinician) and the point selected by our algorithm.

Table 4. List of discriminators tested in the second experiment

1	HS histogram formed from the entire image (HS_ent)
2	HS histograms formed from 1) the entire image and 2) SubIR (classifies SubIR images) + HS histogram formed from the entire image (classifies the remaining images)(HS_ent_reg__HS_ent)
3	3D LBP histogram formed from the entire image (LBP_ent)
4	3D LBP histograms formed from 1) the entire image and 2) SubIR + 3D LBP histogram formed from the entire image (LBP_ent_reg__LBP_ent)
5	3D LBP and HS histograms formed from the entire image (LBP_ent_HS_ent)
6	3D LBP and HS histograms formed from 1) the entire image and 2) SubIR + 3D LBP and HS histograms formed from the entire image (LBP_ent_reg_HS_ent_reg__LBP_ent_HS_ent)

Table 5 shows the mean and the median error of those discriminators. Results of the discriminator experiments are generally consistent with those obtained for the respective classifiers. It can be clearly seen that the best performance is achieved by the classifiers which combine more features of the whole image and SubIR. As expected the localization of esogastric junction is the most accurate, followed by pylorus. With regard to the IV location, the results deteriorate, but are still satisfactory if we consider that in many cases it is impossible to know exactly when the capsule passes from the intestine to colon due to faeces occlusion.

Table 5. The median and mean errors of HMM discriminators in frames.

Discriminator	Esogastric Junction		Pylorus		Ileocecal Valve	
	median	mean	median	mean	median	mean
HS_ent	13	53	186	1174	503	3363
HS_ent_reg_HS_ent	12	50	173	1051	519	3407
LBP_ent	**10**	62	146	875	1770	3751
LBP_ent_reg_LBP_ent	11	70	124	483	1710	4367
LBP_ent_HS_ent	11	52	132	860	**500**	3583
LBP_ent_reg_HS_ent_reg_LBP_ent_HS_ent	**10**	**43**	**103**	**318**	502	**2951**

6 Conclusions and Acknowledgements

In this work we have developed a hybrid SVC/HMM Wireless Capsule endoscopy video segmentation scheme. We have shown that combining global and regional colour and texture features improves the segmentation results, which could lead to a reduction in the video viewing time.

The authors would like to thank Dr Crawford Jamieson and colleagues at the Norfolk & Norwich University Hospital for the provision of annotated WCE videos and the National Association of Colitis and Crohn's Disease for their financial support towards this research. The authors are particularly thankful to Professor Duncan Bell for his support and assistance throughout this project. Finally, we would like to thank Dr David Connah and Prof. Graham Finlayson for granting us access to their new LBP technique.

References

1. G. Iddan, G. Meron, A. Glukhovsky et al. "Wireless capsule endoscopy." *Nature* **405**, pp. 725–9, 2000.
2. Given Imaging Ltd., Israel, http://www.givenimaging.com.
3. A. Maieron et al. "Multicenter Retrospective Evaluation of Capsule Endoscopy in Clinical Routine." *Endoscopy* **36**, pp. 864–868, 2004.
4. J. Berens, M. Mackiewicz & G. D. Bell. "Stomach, Intestine and Colon tissue discriminators for Wireless Capsule Endoscopy (WCE) images." In *Proceedings of SPIE*, volume 5747, pp. 283–290. 2005.
5. J. Berens, M. Mackiewicz, M. Fisher et al. "Using colour distributions to discriminate tissues in Wireless Capsule Endoscopy images." In *Proceedings of Medical Image Understanding and Analyses 2005 Conference*. Bristol, UK, July 2005.
6. J. Berens. *Image Indexing using Compressed Colour Histograms*. Ph.D. thesis, School of Information Systems, University of East Anglia, Norwich, UK, 2002.
7. M. Coimbra, P. Campos & J. P. S. Cunha. "Extracting clinical information from endoscopic capsules exams using MPEG-7 visual descriptors." In *2nd European Workshop on the Integration of Knowledge Semantic and Digital Media Technologies*. IEE, November 2005.
8. M. Swain & D. Ballard. "Color indexing." *International Journal of Computer Vision* **7(1)**, pp. 11–32, 1991.
9. T. Mäenpää & M. Pietikäinen. "Classification with color and texture: jointly or separately." *Pattern Recognition* **37**, pp. 1629–40, 2004.
10. M. P. T. Ojala & T. Mäenpää. "Multiresolution grey-scale and rotation invariant texture classification with local binary patterns." *IEEE Trans. Pattern Anal. Machine Intell.* **24(7)**, pp. 971–987, 2002.
11. D. Connah & G. D. Finlayson. "Using Local Binary Pattern operators for colour constant image indexing." To appear in CGIV 2006 Conference Proceedings.
12. V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, N.Y., 1995.
13. J. C. Platt. *Advances in Large Margin Classifiers*, chapter Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. MIT Press, 1999.
14. H. Lin, C. Lin & R. Weng. "A note on Platt's probabilistic outputs for support vector machines." Technical report, Department of Computer Science and Information Engineering, National Taiwan University, 2003.
15. T.-F. Wu, C.-J. Lin & R. C. Weng. "Probality Estimates for Multi-class Classification by Pairwise Coupling." *Journal of Machine Learning Reasearch* **5**, pp. 975–1005, 2004.
16. L. R. Rabiner. "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." In *Proceedings of the IEEE*, volume 77. 1989.
17. J. S. Boreczky & L. D. Wilcox. "A Hidden Markov Model framework for Video Segmentation Using Audio and Image Features." In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 6. 1998.

Canonical correlation analysis of sub-cortical brain structures using non-rigid registration

Anil Rao[a] and Kola Babalola[b] and Daniel Rueckert[a]

[a]Department of Computing, Imperial College London
[b]Department of Imaging Science and Biomedical Engineering, University of Manchester

Abstract. In this paper, we present the application of canonical correlation analysis to investigate how the shapes of different structures within the brain vary statistically relative to each other. Canonical correlation analysis is a multivariate statistical technique which extracts and quantifies correlated behaviour between two sets of vector variables. Firstly, we perform non-rigid image registration of 93 sets of 3D MR images to build sets of surfaces and correspondences for sub-cortical structures in the brain. Canonical correlation analysis is then used to extract and quantify correlated behaviour in the shapes of each pair of surfaces. The results show that correlations are strongest between neighbouring structures and reveal symmetry in the correlation strengths for the left and right sides of the brain.

1 Introduction

The area of computational anatomy is a rapidly developing discipline [14]. With the increasing resolution of anatomical scans of the human brain, a number of computational approaches for characterising differences in the shape and neuro-anatomical configuration of different brains have emerged. Morphometric techniques can be classified into techniques that deal with differences in brain shape (deformation-based morphometry [3, 9]) and those which deal with differences in the local composition of brain tissue after removing global shape differences (voxel-based morphometry [1]). Even though both approaches require warping of images into a standard reference space using either elastic [4, 13, 20] or fluid [6–8] registration techniques, they differ fundamentally in the way the resulting deformation fields are used. In deformation-based morphometry the deformation fields themselves are used to study similarities and differences, while in voxel-based morphometry these fields are used principally for normalisation. There is currently an active discussion in the neuroscience community regarding the advantages and disadvantages of both methods [2, 5].

A prominent example of modelling the variability of neuro-anatomical structures across a population is the probabilistic atlas of the human brain developed at the Montreal Neurological Institute (MNI) [10] where MR images from 305 subjects were mapped into stereotactic space, intensity normalised and averaged on a voxel-by-voxel basis as part of the International Consortium for Brain Mapping (ICBM) [19]. An alternative approach uses statistical models such as active shape models [12] or active appearance models [11] to represent the variability of the anatomy across subjects. Statistical shape modelling refers to the analysis of the shapes of sub-structures (such as the lateral ventricles in the centre of the brain) and aims to describe their variation across subjects and between groups of subjects (e.g., comparing ventricle size and shape between Alzheimer's sufferers and age-matched normals). Work on shape modelling is limited by the generally unsolved problems of how to segment these structures and determine correspondences across subjects, and also by the relatively unexplored area of how different structures vary statistically relative to each other within the brain. The latter problem will be addressed specifically in this paper.

In this paper we describe the application of canonical correlation analysis for the analysis of the inter-structure shape variation within the brain. Canonical correlation analysis (CCA) is a multivariate statistical tool for describing and quantifying correlated variation between sets of vector variables. It is an extension of multilinear regression and has been used to analyse data in a number of different application areas. Within the field of imaging, canonical correlation analysis has been previously used in image segmentation of magnetic resonance spectroscopic images [16] and the identification of noise in functional magnetic resonance images [21]. Canonical correlation analysis has also been used to estimate the shapes of obscured anatomical sections of the brain from visible structures in magnetic resonance images [17]. However, there it was used as a predictive tool for a limited number of structures within the brain. Here we use it to extract highly correlated factors (or modes) of variation in shape between a number of different anatomical structures within the brain and an associated correlation coefficient that quantifies the degree of correlation in this shape variation. This reveals statistical dependencies between different shapes in the brain that ultimately we would like to incorporate into a hierarchical model-fitting scheme. In the next section we describe the mathematical formulation of canonical correlation analysis before presenting results of the application of this technique in the analysis and prediction of brain structures in section 3.

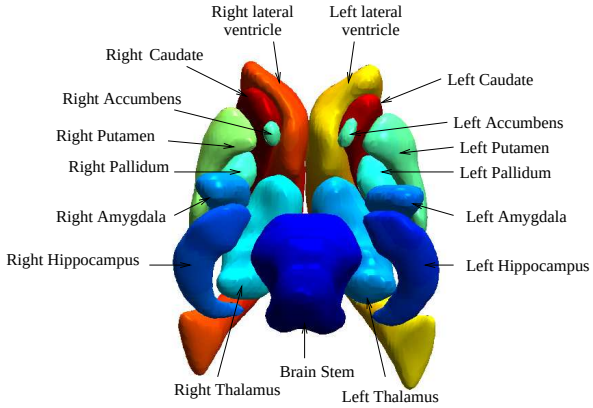


Figure 1. This colour figure shows each of the 17 structures that we considered.

2 Canonical Correlation Analysis

The object of canonical correlation analysis is to extract and quantify correlations between two sets of vector variables, $\mathbf{X} = \{\boldsymbol{x}_i\}, \mathbf{Y} = \{\boldsymbol{y}_i\}$. The technique determines linear combinations of the components of the vector variables in $\mathbf{X}$ that are maximally correlated with linear combinations of the components in $\mathbf{Y}$, and the strength of each of the correlations is described by a corresponding correlation coefficient that lies between zero and one. The linear combinations, known as the canonical modes, give insight into the relationships between the two sets of variables [18].

The canonical modes $\hat{\boldsymbol{a}}_k, \hat{\boldsymbol{b}}_k$ and correlation coefficients ρ_k for $\mathbf{X}$ and $\mathbf{Y}$ are calculated by solving the eigenvalue equations

$$\begin{aligned} C_{XX}^{-1} C_{XY} C_{YY}^{-1} C_{XY}^{T} \hat{\boldsymbol{a}}_k &= {\rho_k}^2 \hat{\boldsymbol{a}}_k \\ C_{YY}^{-1} C_{XY}^{T} C_{XX}^{-1} C_{XY} \hat{\boldsymbol{b}}_k &= {\rho_k}^2 \hat{\boldsymbol{b}}_k \end{aligned} \tag{1}$$

where C_{XX}, C_{YY} and C_{XY} are the covariance matrices describing variation within $\mathbf{X}$, $\mathbf{Y}$, and between $\mathbf{X}$ and $\mathbf{Y}$ respectively [16]. The calculated modes $\hat{\boldsymbol{a}}_k$ and $\hat{\boldsymbol{b}}_k$ are then the linear combinations of variable components in $\mathbf{X}$ and $\mathbf{Y}$ respectively that have a corresponding correlation coefficient ρ_k. The number of modes and correlation coefficients determined by a canonical correlation analysis of $\mathbf{X}$ and $\mathbf{Y}$ will be equal to the minimum of the number of dimensions in the vectors $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$. A single correlation coefficient ρ representing the overall correlation between $\mathbf{X}$ and $\mathbf{Y}$ can be determined by averaging the correlation coefficients over all calculated canonical modes.

Canonical correlation analysis has certain maximal properties similar to those of principal components analysis (PCA). However, the two techniques differ fundamentally in that while CCA focuses on relationships *between* two groups of variables, PCA considers interrelationships *within* a single group of variables [18]. If, for example, we were to pool the two sets of variables $\mathbf{X} = \{\boldsymbol{x}_i\}$ and $\mathbf{Y} = \{\boldsymbol{y}_i\}$ into a single set and then perform a PCA, we would lose the distinction between the two sets of data as the PCA does not 'know' to which data set each variable originated from. The resulting modes would then just model the variation of the composite data set without explicitly describing the dependencies of the individual data sets on each another.

3 Canonical Correlation Analysis of Brain MR Data

3.1 Method

A set of MR images of 93 subjects from the Centre for Morphometric Analysis (CMA), Boston, was used to create a training set of surfaces over which the canonical correlation analysis was applied. The images were obtained at resolution 1mm x 1.5mm x 1mm and had been manually labelled in order to delineate structures within the brain by experts at the CMA. Firstly, a reference subject was chosen and the surfaces of 17 different sub-cortical brain structures of this reference subject were calculated from its labelled image. These surfaces are shown in figure 1 and represent the left and right lateral ventricle, left and right caudate, left and right putamen, left and right accumbens, left and right pallidum, left and right thalamus, left and right amygdala, left and right hippocampus and the brain stem.

In order to model the variation in the surfaces of these structures across all subjects, correspondences between each reference surface point and the corresponding surface in each of the other subjects must be determined. These were cal-

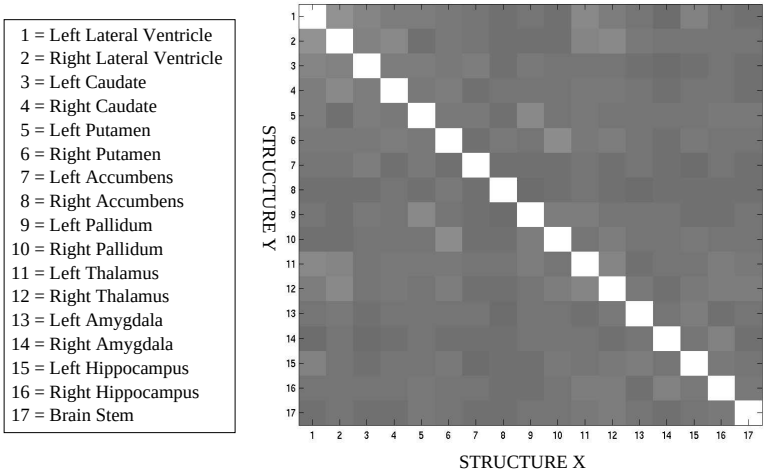


Figure 2. This figure shows the matrix image of the correlations ρ between each pair of structures. In the image, bright areas represent strong correlations close to 1, while dark areas represent weaker correlations close to 0.5

culated by registering the labelled magnetic resonance images of each subject to the reference image using a B-Spline FFD registration algorithm which represents each transformation as a sum of a global affine component and a non-rigid component based on B-Splines [20]. The optimal transformation is found by maximising the label consistency of the labelled images which measures the degree of their alignment.

The registrations were then used to create a set of surface points for each of the structures over all 93 subjects. Firstly, for a given non-reference subject, the corresponding reference to subject transformation was applied to the reference surfaces to give a set of subject surface points. Then, for each structure, the calculated surface points of the 93 subjects were Procustes-aligned using scaling, rotations and translations [12, 15] to give the final surface point coordinates for each structure and subject. The Procustes-alignment ensures that any subsequent statistical modelling of the generated surface points only describes variations in the shape of a structure over the training data rather than variation due to differences in the position, orientation or size of a structure.

For each individual structure X, the vectors $\boldsymbol{x}_i$ representing the surface point coordinates for structure X of the ith subject $0 < i < 93$, were pooled to form a set of vectors $\mathbf{X} = \{\boldsymbol{x}_i\}$. Prior to performing the canonical correlation analysis of surface point coordinates, a principal components analysis was performed on the surface point coordinates for each individual structure across the training data to reduce the dimensionality of the data. The dimension reduction minimizes the computational memory burdens of the canonical correlation analysis and also eliminates colinearity in the data which can cause instability in the calculation of CCA. Fifty-five modes of variation were retained from the prinicipal components analysis of each structure ensuring that at least 95% of the variation in that structure across the training data could be represented. For each structure X, the set $\mathbf{X} = \{\boldsymbol{x}_i\}$ was then transformed into its corresponding principal components basis to give a new set of vectors $\tilde{\mathbf{X}} = \{\tilde{\boldsymbol{x}}_i\}$. Canonical correlation analysis was then performed for all pairs of structures X and Y using the corresponding sets of vectors $\tilde{\mathbf{X}} = \{\tilde{\boldsymbol{x}}_i\}$ and $\tilde{\mathbf{Y}} = \{\tilde{\boldsymbol{y}}_i\}$. In each case, fifty-five canonical modes and correlations are determined describing the correlated behaviour between structure X and structure Y. For a given pair of structures X and Y, these correlations were then averaged to give a final correlation coefficient ρ between zero and one describing the strength of the correlations between the two structures. We chose not to retain the actual canonical modes as we were interested in quantifying correlations rather than analyzing the correlated behaviour itself.

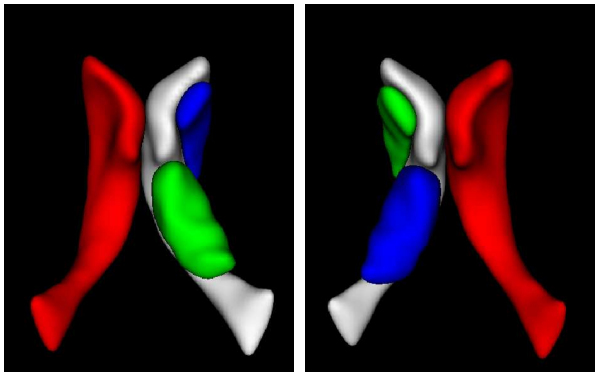
3.2 Results

Figure 2 shows the canonical correlation coefficients for each pair of structures as a grey-level matrix image in which brighter areas represent higher correlations. We can see that the correlation coefficients achieve the maximum value of one along the top-left to bottom-right diagonal as this line represents the correlations between a structure and itself which are always perfect. The matrix image is also symmetrical about this diagonal as the calculation of the correlation coefficients between any two structures is independent of which is taken to be structure X and which is taken to be structure Y.

Structure	Most Correlated Structure	ρ	Structure	Most Correlated Structure	ρ
L. Lateral Ventricle	R. Lateral Ventricle	0.7836	L. Pallidum	L. Putamen	0.7714
R. Lateral Ventricle	L. Lateral Ventricle	0.7836	R. Pallidum	R. Putamen	0.7748
L. Caudate	L. Lateral Ventricle	0.7604	L. Thalamus	L. Lateral Ventricle	0.7659
R. Caudate	R. Lateral Ventricle	0.7703	R. Thalamus	R. Lateral Ventricle	0.7681
L. Putamen	L. Pallidum	0.7714	L. Amygdala	L. Hippocampus	0.7492
R. Putamen	R. Pallidum	0.7748	R. Amygdala	R. Hippocampus	0.7509
L. Accumbens	L. Caudate	0.7437	L. Hippocampus	L. Lateral Ventricle	0.7509
R. Accumbens	R. Putamen	0.7406	R. Hippocampus	R. Amygdala	0.7509
Brain Stem	R. Thalamus	0.7483			

Table 1. This table lists each structure and its most correlated structure

If we consider those correlations lying off the leading diagonal we can see that each pair of structures is correlated to different degrees. For example, the right pallidum is better correlated to the right putamen ($\rho = 0.7748$) than to the left accumbens ($\rho = 0.7236$). This means that, across the training data, the shape of the right pallidum varies in a more correlated fashion with the shape of the right putamen than with the shape of the left accumbens. In table 1 we show the best correlated structure (strongest correlate) for each of the 17 structures and we can see that the strongest correlates of each structure are proximal to it. For example, the left/right caudates are most strongly correlated with the left/right lateral ventricles respectively and we can see in figure 1 that these structures are next to each other. Similarly, each of the lateral ventricles, putamen, accumbens, pallidum, thalamus and amygdala are best correlated with proximal structures. Intuitively, this makes sense as one would expect variation in the shape of a structure to be reflected in the shapes of proximal and neighbouring structures. An example of this relationship between proximity and correlation strength is visualised in figure 3, in which the three strongest correlates for the left/right lateral ventricles are shown. In this figure the lateral ventricles are shown in grey and the correlated structures are shown in red, green and blue in order of decreasing correlation strength. There is also a degree of symmetry in the strongest correlates for structures



(a) (b)

Figure 3. This figure shows each of the lateral ventricles in grey with the three most strongly correlated structures to each of them in in red, green and blue in order of decreasing correlation strength. Figure (a) shows the left lateral ventricle (grey), right lateral ventricle (red), left thalamus (green) and left caudate (blue) and figure (b) shows the right lateral ventricle (grey),the left lateral ventricle (red), the right caudate (green) and the right thalamus (blue).

that appear in both sides of the brain. For example, the left putamen is most strongly correlated with the left pallidum, while conversely the *right* putamen is most strongly correlated with the *right* pallidum. This symmetry is repeated for all the structures apart from the hippocampus and accumbens.

In order to investigate correlations that are not associated with symmetries in the brain, we also performed a canonical correlation analysis in which equivalent structures on each side of the brain were concatenated into single structures. The resulting surfaces were Procustes aligned and 56 principal components of the PCA were retained before performing the canonical correlation analysis.

4 Discussion

The results of the canonical correlation analysis imply that there are differing degrees of correlated variation between the shapes of different structures within the brain. The CCA gives larger correlations for structures that are close to each other, which suggests that the shapes of structures both influence, and are influenced by, those structures that are proximal to them. Intuitively, this makes sense as one would expect variation in the shape of a structure to be reflected in the shapes of proximal and neighbouring structures. However, it should be noted that the correlations and associated modes determined by canonical correlation analysis do not necessarily describe a large amount of the *variation* between structures but instead describe the most *correlated* behaviour in that variation. The most interesting aspect of this work that we are currently pursuing is the incorporation of canonical correlation analysis into a hierarchical model-fitting algorithm. Since CCA quantifies the strengths of the correlations between the shapes of different structures, it can be combined with shape prediction techniques such as partial least squares regression to guide a model fitting in a hierarchical fashion. Such a technique would involve performing CCA on individual structures and groups of structures as we have described in this paper.

References

1. J. Ashburner and K. J. Friston, *Voxel-based morphometry – the methods*, NeuroImage **11** (2000), no. 6, 805–821.
2. ———, *Why voxel-based morphometry should be used*, NeuroImage **14** (2001), no. 6, 1238–1243.
3. J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, and K. Friston, *Identifying global anatomical differences: Deformation-based morphometry*, Human Brain Mapping **6** (1998), 638–57.
4. R. Bajcsy and S. Kovačič, *Multiresolution elastic matching*, Computer Vision, Graphics and Image Processing **46** (1989), 1–21.
5. F. L. Bookstein, *"Voxel-based morphometry" should not be used with imperfectly registered images*, NeuroImage **14** (2001), no. 6, 1452–1462.
6. M. Bro-Nielsen and C. Gramkow, *Fast fluid registration of medical images*, Proc. 4th International Conference Visualization in Biomedical Computing (VBC'96), 1996, pp. 267–276.
7. G. E. Christensen, S. C. Joshi, and M. I. Miller, *Individualizing anatomical atlases of the head*, Proc. 4th International Conference Visualization in Biomedical Computing (VBC'96), 1996, pp. 434–348.
8. G. E. Christensen, M. I. Miller, J. L. Mars, and M. W. Vannier, *Automatic analysis of medical images using a deformable textbook*, Computer Assisted Radiology (Berlin, Germany), Springer-Verlag, June 1995, pp. 146–151.
9. M. K. Chung, K. J. Worsley, T. Paus, D. L. Collins C. Cherif, J. N. Giedd, J. L. Rapoport, and A. C. Evans, *A unified statistical approach to deformation-based morphometry*, NeuroImage **14** (2001), no. 3, 595–606.
10. D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, *Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space*, Journal of Computer Assisted Tomography **18** (1994), no. 2, 192–205.
11. T. F. Cootes, G. J. Edwards, and C. J. Taylor, *Active appearance models*, Proc. 5th European Conference on Computer Vision (ECCV'98), vol. 2, 1998, pp. 484–498.
12. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, *Active Shape Models - their training and application*, Computer Vision and Image Understanding **61** (1995), no. 1, 38–59.
13. J. Gee, M. Reivich, and R. Bajcsy, *Elastically deforming 3D atlas to match anatomical brain images*, Journal of Computer Assisted Tomography **17** (1993), no. 2, 225–236.
14. U. Grenander and M. I. Miller, *Computational anatomy: An emerging discipline*, Quarterly of Applied Mathematics **56** (1998), no. 4, 617–694.
15. B. Horn, *Closed-form solution of absolute orientation using unit quaternions*, Journal of the Optical Society of America **4** (1987), 629–642.
16. T. Laudadio, P. Pels, L. Lathauwer, P. Hecke, and S. Huffel, *Tissue segmentation and classification of mrsi data using canonical correlation analysis*, Magnetic Resonance in Medicine **54** (2005), 1519–1529.
17. T. Liu, D. Shen, and C. Davatzikos, *Predictive modeling of anatomic structures using canonical correlation analysis*, IEEE International Symposium on Biomedical Imaging, 2004.
18. K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate analysis*, Academic Press, Belfast, U. K., 1982.
19. J. Mazziotta, A. Toga, A. Evans, P. Fox, and J. Lancaster, *A probabilistic atlas of the human brain: Theory and rationale for its developement. the international consortium for brain mapping*, NeuroImage **2** (1995), no. 2, 89–101.
20. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, *Non-rigid registration using free-form deformations: Application to breast MR images*, IEEE Transactions on Medical Imaging **18** (1999), no. 8, 712–721.
21. L. Zollei, L. Panych, E. Grimson, and W. Wells, *Exploratory identification of cardiac noise in fmri images*, Sixth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '03), Lecture Notes in Computer Science, 2003.

Consistent 3D Myocardial Contractility Mapping with Variational Restoration

Andrew Huntbatch, Su-Lin Lee, David Firmin and Guang-Zhong Yang.

Royal Society/Wolfson MIC Laboratory, Department of Computing,
Imperial College London, United Kingdom.

Abstract. The study of myocardial mechanics is an important step in the diagnosis and assessment of myocardial dysfunction. The use of phase contrast MR velocity mapping can provide detailed information for revealing intrinsic myocardial contractility. Direct derivation of myocardial contractility based on the acquired velocity data, however, is currently prohibited by the relatively low signal-to-noise ratio because contractile indices such as strain-rate require spatial derivation which can further amplify the level of noise involved. This paper investigates the use of total variation restoration based on the first-order Lagrangian method for improving the internal consistency of myocardial velocity mapping. It is demonstrated that the use of total variation restoration combined with improved MR sequence design to alleviate phase errors and respiratory induced cardiac deformation offers great potential in using myocardial velocity mapping as a routine clinical tool.

1 Introduction

Traditional cardiac MRI approaches for measuring contractility rely on tagging of the myocardium with fiducial markers and require a lengthy and often subjective dependant post-processing procedure. To elucidate intrinsic mechanics related to myocardial dysfunction, MR phase contrast velocity mapping is emerging as a valuable tool clinically because of the detailed contractility information it provides, which permits accurate and reproducible quantification of cardiac structure and function. One of the hurdles in myocardial velocity mapping is in MR sequence design where flow induced phase errors can significantly affect the consistency of the data. Recently, the use of effective blood saturation combined with retrospective respiratory motion management has permitted the acquisition of 3D myocardial velocity data throughout the cardiac cycle. Direct derivation of myocardial contractility based on the acquired velocity data, however, can be affected by the relatively low signal-to-noise ratio. This is particularly problematic for using contractile indices such as strain-rate where the inherent spatial derivation process can further amplify the level of noise involved.

Recently, extensive work has been conducted in velocity field restoration techniques due to the wide range of applications including colour image restoration, optical flow analysis, and diffusion tensor imaging. The most important factor when selecting the restoration method is its ability to distinguish between noise and actual discontinuities in the underlying data. The use of Total Variation (TV) restoration has been proven to be an effective method for cardiac flow analysis [1]. Like the majority of restoration techniques, the TV method is a functional minimization problem with constraints. In addition, the use of the TV norm means that discontinuities are not penalized [2]. Previous work has already applied TV-restoration methods to problems such as restoration of non-flat image features [3] and colour images [4]. The purpose of this paper is to demonstrate how the TV method can be applied to noisy phase contrast myocardial velocity data to derive consistent results for myocardial contractility analysis. The restoration technique employed in this study is based on the TV method first proposed by Rudin [2] and then later extended with the First Order Lagrangian method [1].

2 Method

2.1 Data Acquisition

Seven normal subjects were scanned using a three directional velocity mapping sequence with free-breathing on a 1.5T Siemens Sonata MRI scanner. Cardiac deformation due to respiration was accounted for by using diaphragmatic navigator echoes positioned through the dome of the diaphragm. Images were acquired using a gradient-echo phase-contrast sequence (TR = 53ms, TE = 7.1ms, in-plane resolution = 1.17×1.17mm, field of view of 30×30cm). The bio-polar velocity encoding gradient was set to encode a velocity range of $\pm 15 cms^{-1}$. In total, three phase-contrast images were acquired at each slice for each time frame, as well as one reference phase image used for background phase subtraction to remove phase errors due to susceptibility and eddy currents. These slices were positioned across the short-axis at 8mm intervals from the apex to the base of the heart. Pilot long-axis images were

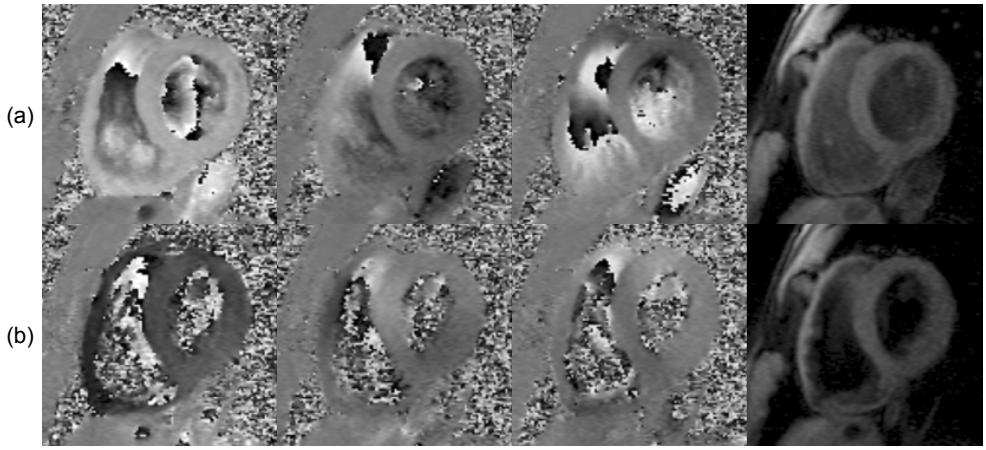


Fig. 1. Example phase contrast myocardial velocity images acquired at mid-systole (a) and mid-diastole (b), showing the velocity components along the *z*, *x*, and *y* directions, as well as the corresponding MR magnitude images.

acquired at the beginning of the sequence for later use in locating the valve planes. 3D velocity fields for each slice were generated by combining the three phase contrast images at each slice. Wraparound was observed in several subjects during peak contraction/relaxation and a phase unwrapping technique was used to correct for this phase aliasing effect [5].

2.2 TV Restoration

Following phase data pre-processing, the TV technique was used to restore the 3D vector data. The basis of this method is to solve a constrained optimization problem. In this case, the property being optimized is the smoothness or variance of the image and the constraint is that the total amount of variance removed from the image is equal to the estimated variance of the noise component. The corresponding Lagrange function *L* can be described as follows:

$$L(u;\lambda) = \sum_{\alpha}\left[\sum_{\beta\in N\alpha} d_l^2(u_\beta, u_\alpha)\right]^{\frac{1}{2}} + \frac{\lambda}{2}\left[\sum_{\alpha} d_l^2(u_\alpha, u_\alpha^0) - |\Omega|\sigma^2\right] \tag{1}$$

where λ is the Lagrange multiplier, d is the embedded Euclidean distance function, u is the image, α and β are pixel locations and Ω is the image domain. The first-order Lagrangian method is employed to guarantee convergence to the optimal solution. The variance of the noise was estimated from the original phase data, but it has been shown that the result is not significantly dependent on this value [1]. The method produces an iterative system to converge on the solution. The equations for this system are as follows:

$$u_\alpha^{n+1} = u_\alpha^n + \Delta t \cdot \prod_{u\alpha}\left[\sum_{\beta\in N\alpha} wt_\alpha^\beta u_\beta + \lambda^n u_\alpha^0\right]$$

$$\lambda^{n+1} = \lambda^n + \Delta t \cdot \frac{1}{2}\left[\sum_{\alpha\in\Omega} d_l^2(u_\alpha, u_\alpha^0) - |\Omega|\sigma^2\right] \tag{2}$$

where

$$wt_\alpha^\beta = \left(\frac{1}{e(u;\alpha)} + \frac{1}{e(u;\beta)}\right)$$

and $e(u;\alpha)$ denotes the strength of voxel α (the sum of the differences between α and its neighbours) and Δt the iteration time-step.

The result of this process is the correction of the directional fields, which removes the effect of noise on the acquired data. As only the velocity directions are rectified by using Eqs (1) and (2) above, further filtering of the velocity magnitude was applied.

2.2 Strain-rate Calculation

After restoration, the model was interpolated to increase the number of short axis slices and therefore increase the density of samples in the longitudinal direction. For every pair of neighbouring real slices, two intermediate slices were generated, by first interpolating the epicardial and endocardial contours linearly and then using these contours to map positions for the velocity data on the new slices. This is then followed by strain-rate calculation. Strain-rates have the ability to provide meaningful contractility information at a voxel level. The strain-rate is the spatial derivative of the velocity data, so for longitudinal strain-rate, as an example, this is the rate of change of the component of the velocity vector in the longitudinal direction over distance. For the myocardium, it is also standard to calculate strain-rates in the radial and circumferential directions. In this study, longitudinal strain rates were calculated at every point on each slice for two cardiac slices: early-systole and early-diastole, which are the phases of peak LV shortening and lengthening respectively. The results were visualized as semi-transparent slices, using a colour ramp to represent positive and negative longitudinal strain-rate. The transparency was decreased proportionally to absolute strain-rate in order to highlight variations.

3 Results

Examples of the phase contrast velocity images used in this study are displayed in Fig. (1), demonstrating the effectiveness of the blood saturation sequence to the improvement of the velocity phase data. The restored longitudinal velocity renderings across six different phases of the cardiac cycle for one of the subjects studied are shown in Fig. (2). Fig. (3) shows the corresponding longitudinal strain-rates derived with and without TV restoration, zoomed in to the mid-ventricular region. It is evident that TV restoration provides a good basis for the subsequent derivation of contractile indices, as the restored images show much less noise and provide a clearer indication of myocardial strain-rate. The standardized AHA 17-segment myocardial model [7] was used to investigate the averaged longitudinal strain-rates. The results for the mid-ventricular region of the seven subjects studied are listed in Fig. 4, which are consistent with other studies [6] depicting strain-rate variation across different phases of the cardiac cycle.

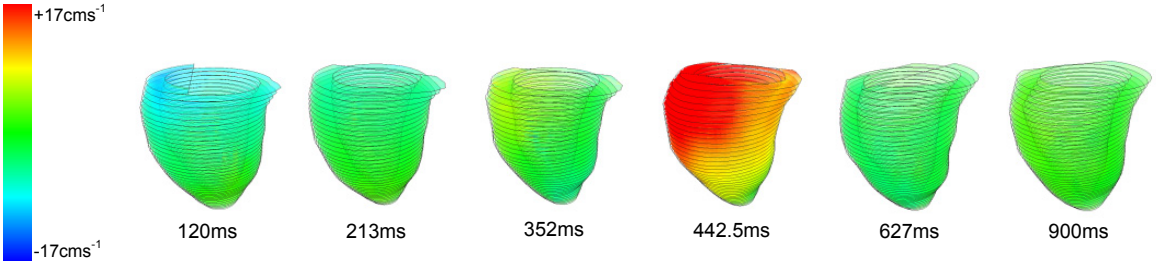


Fig. 2. Restored longitudinal myocardial velocity for one of the subjects studied showing velocity distribution at different R-wave delays in *ms*.

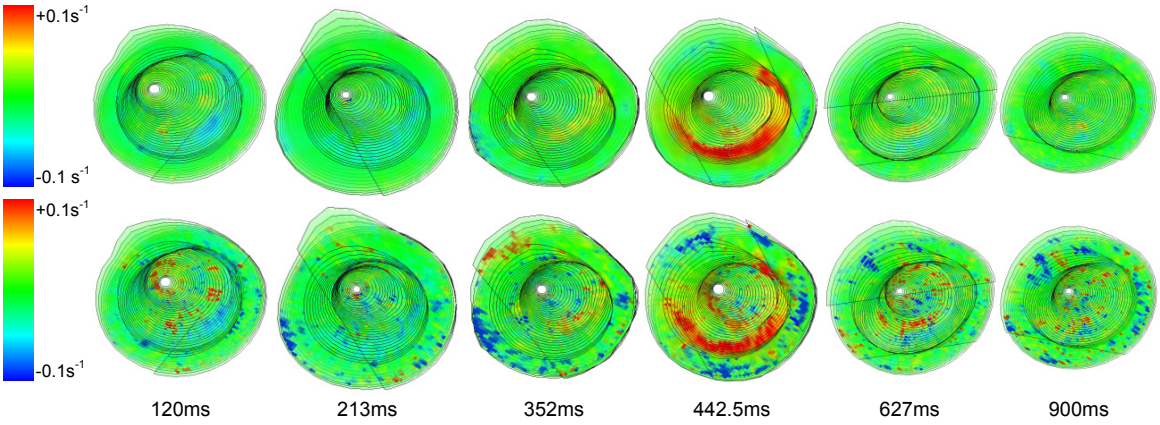


Fig. 3. A comparison of detailed strain-rate distribution corresponding to Fig. 2 with (top) and without (bottom) variational restoration.

4 Discussion and Conclusions

The use of 3D phase contrast data in the derivation of meaningful strain-rates has the potential to provide an efficient method of detecting the presence of myocardial dysfunction. Previous work involving strain-rates produces averaged results over large segments which prohibits the observation of detailed contractile patterns. The use of TV restoration combined with improved MR sequence design to alleviate phase errors and respiratory induced cardiac deformation offers great potential in using myocardial velocity mapping as a routine clinical tool.

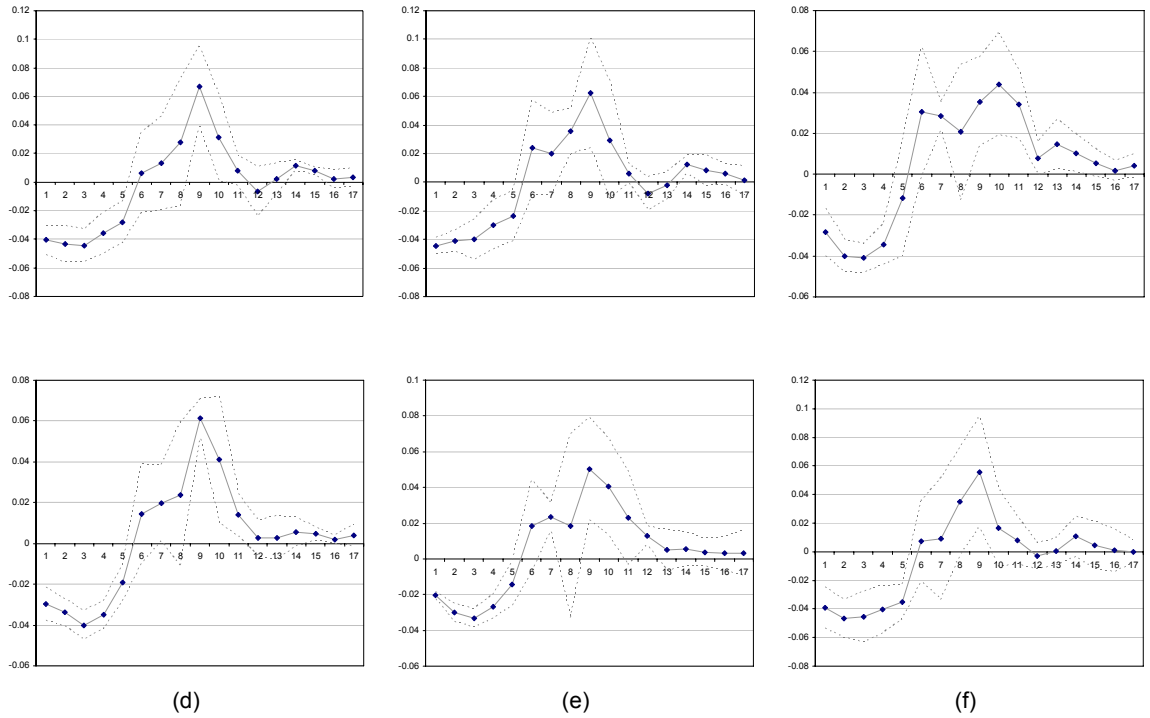


Fig. 4. Mean longitudinal strain-rates (in s^{-1}) against timeframe for six of the mid-ventricular regions of the LV showing the contractile pattern of the a) anterior, b) anterolateral, c) inferolateral, d) inferior, e) inferoseptal and f) anteroseptal regions. The dotted lines indicate the standard deviation.

References

1. Yin-Heung Pauline Ng, Guang-Zhong Yang "Vector-valued image restoration with applications to magnetic resonance velocity imaging", Journal of WSCG Vol. 11, No.1, 2003.
2. Leonid I Rudin, Stanley Osher and Emad Fatemi "Nonlinear total variation based noise removal algorithms", Physica D 60, pp 259-268, 1992.
3. Tony Chan and Jianhong Shen "Variational Restoration of Nonflat Image Features: Models and algorithms", SIAM Journal of Applied Mathematics 61, No. 4, pp 1338-1361, 2000.
4. Peter Blomgren and Tony Chan "Color TV: Total Variation Methods for Restoration of Vector Valued Images", IEE Transactions on Image Processing 7(3), pp 304-309, 1998.
5. Guang-Zhong Yang, Peter Burger, Philip Kilner *et al.* "Dynamic range extension of cine velocity measurements using motion registered spatiotemporal phase unwrapping", Journal of Magnetic Resonance Imaging 6, issue 3, pp 495-502, 1996.
6. Steffen Petersen, Bernd Jung, Frank Wiesmann *et al*. "Myocardial Tissue Phase Mapping with Cine Phase-Contrast MR Imaging: Regional Wall Motion Analysis in Healthy Volunteers", Radiology 238, No. 3, pp 816-826, 2006.
7. Manuel Cerqueira, Neil Weissman, Vasken Dilsizian *et al*. "Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart", American Heart Association Circulation 105, pp 539-542, 2002.

Quantification of attenuation-correction artefacts in cardiac PET/CT caused by respiratory motion

Sarah J. Martin*, Brian F. Hutton

Institute of Nuclear Medicine, University College London, 235 Euston Road (T5), NW1 2BU, UK

Abstract. The effects of attenuation mismatches on cardiac PET/CT studies were investigated using the NCAT software phantom. Different respiratory states were simulated and projections of the activity and attenuation maps were formed. A mismatch in the attenuation was simulated by correcting the PET projections with an incorrect attenuation map, prior to reconstruction. It was found that attenuation mismatches can produce severe artefacts in reconstructed images, leading to an apparent variation in myocardial activity over the left ventricle. The variation in counts in an image with matched attenuation (and therefore error free) was found to be 4.9% of the mean when averaged over 10 regions of interest. The most prominent source of artefact was found to be attenuation mismatches due to cardiac translation during respiration, which caused the variation to rise to 15.6%. Mismatches due to changing lung volume (where the lung stretches in the superior-inferior direction due to diaphragmatic contraction) also caused artefacts, increasing the variation to 6.8% of the mean. Radial stretching of the lung however had a negligible effect. Motion correction algorithms should therefore concentrate primarily on the correct measurement of cardiac position, but the effect of changing lung geometry due to diaphragmatic contraction should also be considered.

1 Introduction

Positron Emission Tomography (PET) can be used to assess myocardial perfusion by imaging the distribution of an injected radionuclide tracer in the left ventricular myocardium. This technique has the ability to generate quantitative measurements of regional perfusion provided that the necessary corrections are applied (for example, attenuation correction) and that the presence of motion does not degrade the image quality. Excluding the presence of motion however is not trivial due to the long scan time (typically 3-5 minutes) and hence the unavoidable inclusion of respiratory motion. The PET data therefore represent an average over the respiratory cycle.

With the advent of combined PET/CT scanners, it is now commonplace to use CT data to derive attenuation-correction factors for correcting the PET data. This poses an additional problem, due to the different scanning techniques employed by the two modalities and the resulting mismatches between them that may occur. Rather than representing the average over the respiratory cycle, the CT is acquired with a single pass over the patient, and therefore only a particular phase of the respiratory cycle is captured at each point. Therefore, in addition to the blurring of the PET data, there is also the possibility of an erroneous attenuation correction being applied, causing a degradation of quantitative accuracy. An accurate attenuation correction is an important stage in producing quantitatively correct PET images and mismatches often lead to artefacts appearing in areas such as the lower lung and liver [1-3]. Due to the heart being positioned in close proximity to the diaphragm in the lower lung, it is particularly sensitive to the effects of respiratory motion [4, 5] and hence measures need to be taken to prevent the occurrence of these artefacts.

The aim of this study was to quantify the changes produced by attenuation mismatches, which has applications in the development of a motion-correction technique. Due to the complex changes in geometry in the region of the heart during respiration, there are a number of factors that can contribute to attenuation mismatches, for example the motion of the heart, the changing geometry and density of the surrounding lung tissue and the motion of the diaphragm. It is important therefore to identify the key factors that lead to the production of artefacts, for the purpose of being able to construct a suitable motion-correction technique.

2 Method

2.1 The NCAT phantom – generation of activity and attenuation maps

The NCAT (NURBS based cardiac-torso) software phantom [6] was used to generate activity and attenuation maps at different respiratory states. This phantom simulates the anatomy of a patient as it changes over the respiratory cycle and is based on the Visible Human dataset [7], which was developed from CT and MRI datasets of normal subjects. It allows investigation of differing breathing patterns on observed activity distributions and has been widely used in Nuclear Medicine for both cardiac and whole-body PET studies [1, 8, 9]. The phantom allows the

* Contact: sarah.martin@uclh.nhs.uk

user to specify parameters such as the activity distribution, the motion present (respiratory, cardiac or both), as well as parameters relating to the nature of the motion. For respiratory motion, this includes the period of the motion and the extent of the lung expansion, which is controlled by the amplitudes of the diaphragm and chest wall motions. The motion of the heart during respiration in the phantom is primarily a translation in the cranio-caudal direction, the amplitude of which depends on the motion of the diaphragm. In normal breathing, both types of breathing motion (diaphragmatic and chest) contribute approximately equally to the change in volume of the lungs, causing stretching of the lung tissue both radially and in a cranio-caudal direction. However, this varies considerably between patients, with some breathing with primarily diaphragmatic motion, and some with primarily chest wall motion [10]. This is an important consideration in artefact assessment, since the motion of the heart as well as the geometry of the lungs in relation to the heart vary considerably with these two motion types.

Datasets were generated for the purposes of investigating the effect of attenuation mismatches caused by the following factors: a) cardiac motion due to normal respiratory motion b) changing lung geometry due to diaphragmatic contraction and c) changing lung geometry due to chest wall motion. The methods used to create the datasets took into account the processes that were being investigated so that each factor could be isolated. Diaphragmatic contraction causes stretching of the lung tissue primarily in the cranio-caudal direction, such that the lung attenuation in any individual transaxial slice is reduced. This through-plane stretching was investigated therefore by simply altering the value of the lung attenuation coefficient, which simulates the addition of air while excluding any additional effects of motion. Chest wall motion however, causes an in-plane stretching, which results in both a change in the lung attenuation coefficient and a change in geometry within each slice.

For all datasets, only respiratory motion was simulated (i.e. no cardiac motion was involved) and the period of the respiratory cycle was chosen to be 5s. The parameters specifying the extent of the diaphragm and chest wall motions (where present) were set to the default values for normal breathing, as listed in table 1. These default values were originally derived by setting the NCAT parameters such that the resulting lung volume curve matched that of normal breathing [6].

	Baseline (no motion)	All motion involved in respiration	Lung expansion (diaphragm contraction only)	Lung expansion (chest wall motion only)
Dataset name:	baseline	total	lung (diaphragm)	lung (chest)
Amplitude of diaphragm motion (cm): (Normal breathing: 2.0cm)	0.0 cm	2.0 cm	2.0 cm*	0.0 cm
Amplitude of chest wall motion (cm): (Normal breathing: 1.2cm)	0.0 cm	1.2 cm	0.0 cm	1.2 cm
% change in lung volume (inhale to exhale): (Normal breathing: 30%)	0 %	30%	17%	13%

Table 1. NCAT parameters used for each of the datasets generated. * This amplitude was simulated by altering the lung attenuation coefficient values – the NCAT parameter was in fact set to 0.0cm to avoid cardiac translation.

2.2 Image formation

After creating the activity and attenuation maps, images of the simulated data were then formed by performing the following steps: firstly the activity map was projected and attenuated to form a series of sinograms, an attenuation-correction was then applied to these sinograms and finally images were formed using a 2D OSEM reconstruction algorithm. Mismatches between the PET and the CT data were simulated by attenuating the PET data with its matched attenuation map but then performing the attenuation-correction with a different, mismatched attenuation. This allowed the effects of using CT for attenuation-correction to be investigated, by using only a specific phase of the respiratory cycle in the attenuation map. Noiseless, scatter-free images were used and resolution effects were not taken into account, to ensure that the effect of the mismatch was the only factor being investigated. The images that were formed using this technique are as detailed in table 2.

Image name	Activity map used (respiratory phase, dataset name)	Attenuation map used (respiratory phase, dataset name)	Differences between activity and attenuation maps
Baseline	End-exhale, baseline	End exhale, baseline	None – matched attenuation
Total	End-exhale, total	End-inhale, total	Change in diaphragm and chest wall positions and hence different cardiac position and lung volume
Lung (diaphragm)	End-exhale, lung (diaphragm)	End-exhale, lung (diaphragm) (with altered μ values in lung to simulate end-inhale)	Only difference is the μ value for lung (no change in geometry)
Lung (chest)	End-exhale, lung (chest)	End-inhale, lung (chest)	Change in the μ value for lung and the geometry of the lungs (change in chest wall position)

Table 2. Details of the images generated from the NCAT maps (μ is the linear attenuation coefficient)

2.3 Image analysis

The resulting images were assessed in terms of the variation in activity across the myocardium induced by the mismatched attenuation map. The images were reoriented to allow inspection of the horizontal long axis (HLA) and vertical long axis (VLA) of the left ventricle (LV). The reorientation was performed manually by visually aligning the slices with the axes of the heart, which is the method used in clinical practice. Differing cardiac positions in the various datsets prevented the use of a universal rotation and hence this manual method was used. The mid-slices of the HLA and VLA were identified and then used for analysis of the uniformity. These datasets both cut through the centre of the LV and hence both include data at the apex as illustrated in figure 1a. The LV myocardial activity in each of the mid-slices was divided into six regions of interest (ROI) as shown in figure 1b, to obtain measurements from a representative area of the myocardium.

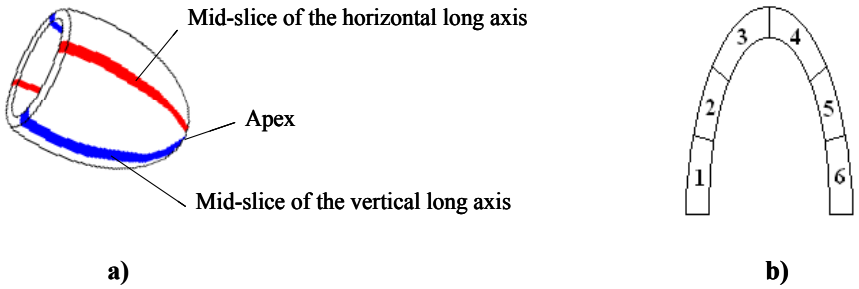


Figure 1. a) Diagram of the left ventricle showing the myocardium visible in the mid-slices of the horizontal and vertical long axes (highlighted). The apex, where the two mid-slices intersect is indicated. b) The six ROIs used for analysis – regions 3 and 4 were excluded from the VLA slice analysis to avoid duplication.

The ROIs were drawn by hand onto one dataset and subsequently copied onto all others for consistency. The average number of counts in each ROI was noted and the standard deviation of these averages was calculated. This method was used in order to minimise the effects of variations perpendicular to the LV wall, while maintaining the ability to measure variations along the myocardium. The standard deviation took into account the averages within only 10 of the ROIs to avoid including the apex data twice, where the HLA and VLA of the left ventricle intersect. The standard deviation as a percentage of the mean was the value used to compare the different images. The value calculated from the baseline image was used as a basis against which to compare the other images, in order to assess the severity of the changes induced by the mismatches.

3 Results and discussion

The mid-slices of the HLA and VLA for each dataset are shown in figures 2 and 3 respectively. The baseline variation obtained with an artefact-free image was found to be 4.9% of the mean when data over 10 ROIs were taken into account. On a visual level, this image appeared to be relatively uniform over the LV myocardium, as expected. When different respiratory states were used for attenuation-correction, artefacts were introduced into the activity distribution, which was reflected in the magnitude of the calculated variance.

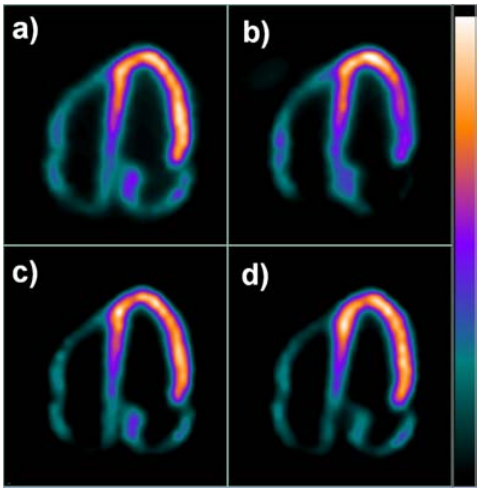


Figure 2. HLA slices for the baseline (artefact-free) image (a) and images containing the following attenuation mismatches: b) cardiac translation, in-slice lung expansion (chest wall motion) and through-slice lung expansion (diaphragm contraction), c) through-slice lung expansion only and d) in-slice lung expansion only.

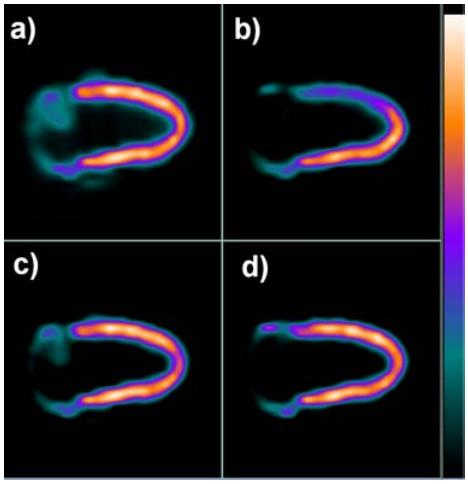


Figure 3. VLA slices for the same datasets as in figure 2.

The maximum variation was found to be 15.6% of the mean, occurring when all sources of attenuation mismatch were included (image b). Visually, this corresponded with severe levels of artefacts that would be diagnostically significant. The variations present within the other sets of data were found to be 6.8% for the through-slice lung expansion (image c) and 5.1% for the in-slice lung expansion (image d). The variation within the in-slice expansion was not noticeably different to that of the error-free image, indicating that this factor does not contribute to the observed artefacts. The through-slice lung expansion variation however, suggested that this does contribute to the production of artefacts if not properly accounted for. This is in agreement with its visual appearance, where a slight reduction in counts can be seen at the apex of the left ventricle. Although it is not possible to isolate the cardiac

translation without a change in the lung volume, the large increase in variation when this factor is included suggests that this is the primary factor leading to artefacts. Moreover, it is possible that the individual factors may oppose each other to some extent, meaning that the increase in variation caused by the change in cardiac position may be even larger than simply the difference between the maximum variance and that caused by the other factors.

4 Conclusions

This study has shown that although the key factor in determining accurate attenuation correction is the correct measurement of the position of the heart, lung geometry factors also need to be considered. As expected, the effect of in-slice lung expansion caused by motion of the chest wall was very small. This is due to the fact that the decrease in attenuation coefficient is counteracted by the increased thickness of lung tissue between the heart and the chest wall. Hence, the overall radiological path-length is not altered and the effect of the mismatch is minimised. When lung expansion occurs in the perpendicular direction however, each transaxial slice is altered and the effects of the mismatch are more prominent. In this case, the apex of the left ventricle was the region most affected, causing the appearance of a slight defect in this area. Although the qualitative changes were fairly subtle, they were measurable on a quantitative level and hence need to be addressed if quantitative accuracy is required.

The aim is to incorporate these findings into the development of a motion-correction technique that will improve the quality of cardiac PET/CT scanning both on a qualitative and a quantitative basis. There are existing techniques that have been developed for this purpose [9], however obtaining an accurate attenuation map in the absence of 4D CT remains a challenge. 4D CT overcomes the problem of finding an accurate attenuation map, however there are high patient doses associated with 4D CT, rendering it undesirable for this purpose. A possible method of avoiding this would be to acquire only a single breath-hold CT and to derive the CT map at the other phases of the respiratory cycle. This would be possible if the registration parameters relating the position of the heart, lungs and diaphragm in each respiratory phase were found from a 4D PET study, either by obtaining the relation between individual PET frames, or by relating each PET frame to the breath-hold CT image. There are further considerations in deciding how to derive these relations and it is proposed that a data-driven model that provides constraints to the registration parameters could be an appropriate method. Once a motion-correction technique has been developed, similar work to that carried out here could be performed for the purposes of assessing the improvements in artefact appearance.

Acknowledgements

This work is supported by funding from BBSRC and GlaxoSmithKline.

References

1. G. M. Fitzpatrick & G. Wells "Mis-registration between emission and transmission data in cardiac PET/CT". In *Nuclear Science Symposium*, pp. 3048-3052, IEEE, 2003.
2. M. M. Osman, C. Cohade, Y. Nakamoto, et al. "Respiratory motion artifacts on PET emission images obtained using CT attenuation correction on PET-CT", *European Journal of Nuclear Medicine and Molecular Imaging* **30**, pp. 603-606, 2003.
3. G. W. Goerres, C. Burger, E. Kamel, et al. "Respiration-induced Attenuation Artifact at PET/CT: Technical Considerations", *Radiology* **226**, pp. 906-910, 2003.
4. K. McLeish, D. L. G. Hill, D. Atkinson, et al. "A study of the motion and deformation of the heart due to respiration", *Medical Imaging, IEEE Transactions on* **21**, pp. 1142-1150, 2002.
5. G. Shechter, C. Ozturk, J. R. Resar, et al. "Respiratory motion of the heart from free breathing coronary angiograms", *Medical Imaging, IEEE Transactions on* **23**, pp. 1046-1056, 2004.
6. W. P. Segars, D. S. Lalush, and B. M. W. Tsui "Modeling respiratory mechanics in the MCAT and spline-based MCAT phantoms", *Nuclear Science, IEEE Transactions on* **48**, pp. 89-97, 2001.
7. M. J. Ackerman "The Visible Human Project", *Proceedings of the IEEE* **86**, pp. 504-511, 1998.
8. F. Lamare, T. Cresson, J. Savean, et al. "Affine transformation of list mode data for respiratory motion correction in PET", pp. 3151-3155, 2004.
9. M. Dawood, N. Lang, J. Xiaoyi, et al. "Lung motion correction on respiratory gated 3-D PET/CT images", *Medical Imaging, IEEE Transactions on* **25**, pp. 476-485, 2006.
10. C. Plathow, H. Zimmermann, C. Fink, et al. "Influence of different breathing maneuvers on internal and external organ motion: Use of fiducial markers in dynamic MRI", *International Journal of Radiation Oncology*Biology*Physics* **62**, pp. 238-245, 2005.

Correcting for Motion between Acquisitions in Diffusion MR Imaging

Yu Bai* and Daniel C. Alexander†

Centre for Medical Image Computing, Department of Computer Science,
University College London, London WC1E 6BT

Abstract. This paper outlines a method for aligning diffusion weighted images in a diffusion tensor MRI sequence. Traditional approaches register all the images to an unweighted image, which can introduce errors particularly for high *b*-value imaging. The new method uses the diffusion-tensor model to predict target images for registration. Experiments show that the new method improves alignment in simulation and removes errors introduced by the traditional approach on a full-brain data set.

1 Introduction

In diffusion-tensor (DT) MRI, a number of diffusion-weighted (DW) images with different diffusion-weighting gradient directions are taken during scanning which can take 20 minutes or more. During such a long time, small head movements are not easy to avoid. However, the tensor calculation assumes that each voxel corresponds to the same anatomical location in all the measurement images. Thus, to fit the diffusion tensor, all the measurements need to align properly. A small bulk motion can cause unmatched measurements to be used during the tensor fitting. The traditional correction approach uses a non-diffusion weighted image i.e. a T2-weighted image as the reference for registration, but the differences between diffusion-weighted images and the non-diffusion weighted reference image cause mismatching to occur during registration. We propose two model-based methods to improve the motion correction and correct the errors that the standard correction scheme introduces. We test the new approaches on both reduced-size data sets with artificial movement corruption and a full data set. Tests on the reduced-size data confirm efficacy of the new method. Tests on the full data set reveal advantages of the new methods over the traditional approach. Section 2 gives background required for the paper. We introduce the new method in section 3. Section 4 shows experiments and results and we conclude in section 5.

2 Background

2.1 Diffusion-Weighted and Diffusion-Tensor MRI

The measurements acquired in DW-MRI are sensitive to the motion of charged particles. The technique allows us to image the diffusion coefficient in biological tissue. The microstructure of biological tissue, such as brain tissue, contains barriers to water mobility that often have preferential orientation, which makes the diffusion coefficient anisotropic. DT MRI [1, 2] measures the apparent diffusion tensor in each voxel by acquiring seven or more DW images with different diffusion-weighting and fitting the tensor in each voxel. This assumes that particle displacements are Gaussian distributed with zero mean and covariance $2tD$ where t is the diffusion time and D is the diffusion tensor. With this model

$$A(\mathbf{q}) = \exp(-t\mathbf{q}^T D\mathbf{q}), \tag{1}$$

where $\mathbf{q}$ is a three vector that depends on the strength and orientation of magnetic-field-gradient pulses in the acquisition and A is the MRI measurement normalized by the measurement with $\mathbf{q} = \mathbf{0}$. The diffusion tensor

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yx} & D_{zz} \end{pmatrix} \tag{2}$$

is a symmetric three-by-three matrix, where D_{xx}, D_{yy} and D_{zz} are diffusion coefficients along x, y and z axes and the D_{xy}, D_{xz} and D_{yz} are correlation coefficients between the axes. The eigenvalues λ_1, λ_2 and λ_3 of D determine the shape of p. The eigenvectors of D, $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$, give the orientation.

To fit the six free parameters in D, a minimum of six measurements $A(\mathbf{q})$ are required with independent $\mathbf{q}$. Most often, the linear least-squares algorithm is used to fit the tensor D to the $\log$ of the measurement via equation (1). However, fitting directly to equation (1) using non-linear optimization can improve accuracy since noise is approximately

*Y.Bai@cs.ucl.ac.uk
†D.Alexander@cs.ucl.ac.uk

Gaussian on the $A(\mathbf{q})$ rather than on $\log A(\mathbf{q})$. Other schemes such as the RESTORE algorithm [3], iteratively reject outlying $A(\mathbf{q})$ during fitting to improve robustness to physiological noise arising from motion or cardiac pulsation.

2.2 Motion Correction

For fitting the diffusion tensor, voxels in different diffusion-weighted images must correspond to the same anatomical location. Thus, all the measurement images need to be well aligned.

As mentioned in the previous section, to fit the six free parameters in D, a minimum of seven measurements are required, which must be acquired from the patient in one scan. In fact, in order to obtain less noisy DT-MR images, around 50 measurements are typically acquired. The whole scan usually lasts around 20 minutes. During such a long time, small head movements, such as translations and rotations are not easy to avoid, especially for baby patients. Furthermore, eddy-current-induced 2-D geometric distortions often remain in diffusion-weighted images. The effect of residual eddy currents is to cause image shearing, scaling and shifting [4].

Image registration warps one image volume X into alignment with another one Y. More precisely, the registration problem is to find the geometric transformation of image Y, that maximizes the similarity between X and Y. To do this, we construct a cost function that quantifies the dissimilarity between two images and search for the transformation T^* which gives the minimum cost [5]:

$$T^* = \arg\min C(Y, T(X)), \tag{3}$$

where $C(I_1, I_2)$ is the cost function, and $T(X)$ is the image X transformed by T. Two of the most commonly used cost functions are normalized correlation (NC) and mutual information (MI).

One of the most common registration problems is the presence of local minima in the cost function, which causes local optimization to fail to find the global minimum. Incorporating a local optimization strategy within a multiresolution framework can solve this problem, which is called 'a global-local hybrid optimization method' by Jenkinson et al. [5]. This means the local optimization works progressively from low resolution images to higher resolution images, which reduces the effects of local minima traps.

Diffusion MRI researchers use registration to align the diffusion-weighted images before fitting the diffusion tensor. Since the T_2-weighted ($b = 0$) image does not suffer from the distortions induced by eddy current and has the highest SNR, it is usually chosen as the reference image for registration [6]. Haselgrove and Moore [6] use cross correlations as the cost function, but when the contrasts of the source and target images are very different, cross correlation performs poorly. Bastin et al. [7] also indicate this approach does not perform well with diffusion-weighted images acquired with b-values higher than $300\mathrm{s\ mm}^{-2}$. Other similar approaches are in [8].

3 Method

We propose a new class of model-based correction methods, specifically FTAM (Fit Tensors to All the Measurements) and FTSM.restore (Fit Tensors to Selected Measurements), to improve the motion correction and remove the errors that the standard correction scheme introduces. Both methods contain four steps, and use different reference images for diffusion-weighted images with different gradient directions for registration, so that the registrations take into account the contrast differences of measurements.

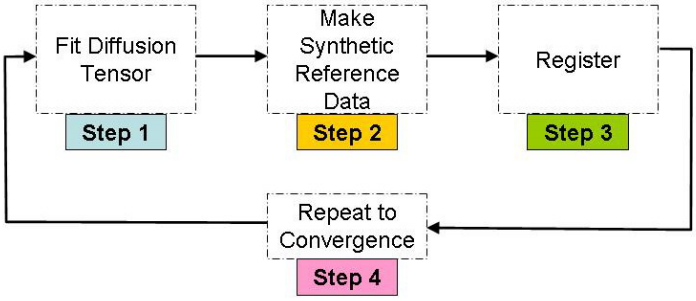


Figure 1. Model Fit Method Flow Chart

Figure 1 shows the basic four-step idea. Step 1 fits the diffusion tensor to the uncorrected data. Step 2 synthesizes reference data for each measurement from the fitted tensor. We use FLIRT [5] to do the registration in Step 3. In Step 4, we repeat the process, regarding the output from Step 3 as the input to Step 1, until convergence, which occurs when the difference S_n between the present output A_{n+1} and the output from previous cycle A_n

$$S_n = \sum_{m=1}^{M} \sum_{i=1}^{V} \sqrt{(A_{n+1}(\mathbf{q}_m)_i - A_n(\mathbf{q}_m)_i)^2}, \tag{4}$$

where M is the number of measurements and V is the number of voxels in every measurement, achieves an acceptably small value.

The only difference between FTAM and FTSM.restore lies in the first step, fitting the diffusion tensor to the measurements. FTAM fits the diffusion tensor D to the whole set of diffusion weighted measurements; FTSM.restore fits the tensor only to a selected subset of measurement images, and it excludes badly misaligned images from the fitting process by using RESTORE outlier-rejection method [3].

The comparison between the traditional correction approach and the model-based method is illustrated in Figure 2.

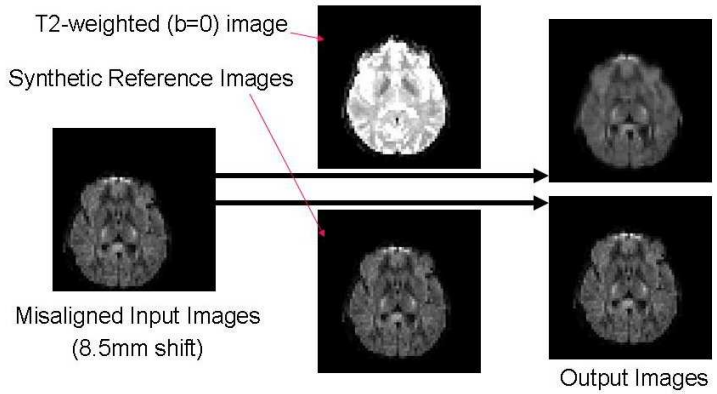


Figure 2. Comparison between Traditional Correction (top) and Model Fit Correction (down)

4 Experiment and Result

The test data set has 66 measurements including 6 with $\mathbf{q} = \mathbf{0}$, with size $128 \times 128 \times 60$ voxels for each measurement. In order to reduce the time required to test and compare the new methods, we have constructed a smaller data set, where we reduce the 3D voxel grid to $64 \times 64 \times 30$, and cut down the measurements number from 66 to 14 which contains two zero-weighted and twelve measurements with non-zero $\mathbf{q}$.

In the experiments to compare the new model-based corrections with the traditional approach, we introduce different levels of artificial shift transformations (i.e. rigid translations) to variable numbers of the measurements. Then, we use Traditional Method, FTAM and FTSM.restore to register the synthetic data sets with different levels of shift corruptions. We use root mean square error, between the registered images and the uncorrupt images, to evaluate the registration results from different methods. The results are summarised in Tables 1, 2, 3 and 4.

Compared to the errors without registration (Table 1), the traditional approach adds extra errors. Although FTAM also adds some errors, it still does much better for smaller corruptions than the traditional approach. The FTSM.restore method is the only one that reduces the errors from the unregistered data. It is evident that both FTAM and FTSM.restore provide better performances than the traditional approach and the FTSM.restore method gives very promising results.

The results in Tables 3 and 4 suggest that FTSM.restore is more robust than FTAM to larger corruptions; FTAM breaks down with smaller shifts or corruption to less images. This is because FTAM fits the tensor to the whole set of diffusion weighted measurements, including badly misaligned images, so that the reference image can be badly corrupted leading to poor realignment particularly for larger displacements. FTSM.restore fits the tensor only to a selected subset of measurement images and it excludes the corrupted part and thus provides more consistent reference images and more reliable realignment. The situation is similar when the number of corrupt measurements increases.

	Corrupt Shift Scale			
Number of Corrupted Measurements	1.7(mm)	3.4(mm)	5.1(mm)	8.5(mm)
1	3.2524	4.5176	5.5744	7.5495
2	6.7285	9.3467	11.457	15.227
3	10.104	13.974	17.145	22.742
5	15.36	21.462	26.326	35.01

Table 1. Root Mean Square Error from Unregistered Data

	Corrupt Shift Scale			
Number of Corrupted Measurements	1.7(mm)	3.4(mm)	5.1(mm)	8.5(mm)
1	28.799	28.794	28.797	28.793
2	28.817	28.795	28.799	28.798
3	28.812	28.804	28.811	28.825
5	28.784	28.835	28.829	28.84

Table 2. Root Mean Square Error from Traditional Method

We know that to fit the six free parameters of the tensor, a minimum of seven measurements are required. Thus, when the number of corrupted measurements is so high that the number of well aligned diffusion-weighted measurements is less than six, even the RESTORE method can not fit the tensor model reliably, which means FTSM.restore does not work very well.

We also run the three methods on the full data set. Figure 3(a) illustrates the misalignment of the corpus callosum outline with the FA map after the standard correction which used the normalized correlation similarity measure to align each diffusion-weighted image to the b=0 image. This is an artifact of the traditional approach that we observe frequently and arises from poor alignment of the DW images with the T_2 image. Figure 3(b) and (c) show that the problem disappears after correction with FTAM and FTSM.restore.

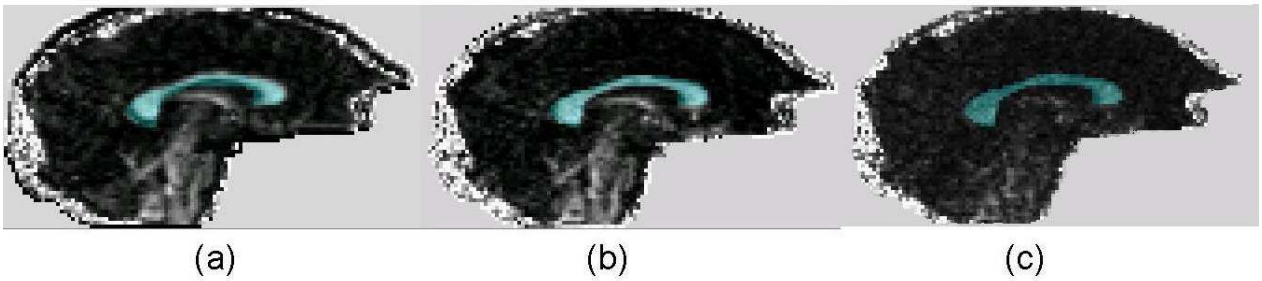


Figure 3. FA maps overlaid with the corpus callosum as segmented by hand from the $b = 0$ image, after (a) standard motion correction, (b) FTAM correction and (c) FTSM.restore.

5 Discussion

The traditional correction approaches use the same reference image to register all the other diffusion-weighted images with different gradient directions. Although they correct both eddy-current-induced distortion and subject motion, the contrast differences still cause misregistration. Our method predicts separate reference images for each diffusion gradient and thus avoids the mismatching caused by the intensity differences between component images. The method should be particularly useful for high b-value imaging and $\mathbf{q}$-space techniques such as [9] and [10] where diffusion-weighted images are very different to the $b = 0$ image.

In this work, we have used only 3D affine transformations to correct for misalignment between acquisitions. In future, to correct for both patient subject motion and eddy-current-induced distortions, we plan to use more flexible transformations such as 3D rigid body combined with slice by slice 2D affine transformations [4].

One potential drawback of not registering directly to the undistorted $b = 0$ image is that the final aligned data set may contain residual distortion. Although that distortion should be consistent among the images, it may cause misalignment

	Corrupt Shift Scale			
Number of Corrupted Measurements	1.7(mm)	3.4(mm)	5.1(mm)	8.5(mm)
1	6.8786	10.1	14.65	21.178
2	10.552	15.583	21.497	32.836
3	12.598	18.948	25.843	39.753
5	15.777	24.325	32.803	48.452

Table 3. Root Mean Square Error from FTAM Method

	Corrupt Shift Scale			
Number of Corrupted Measurements	1.7(mm)	3.4(mm)	5.1(mm)	8.5(mm)
1	2.4611	3.1712	3.318	3.2859
2	4.7569	4.493	4.8102	4.0753
3	5.5724	7.8798	8.7495	8.7861
5	10.589	35.228	17.664	37.127

Table 4. Root Mean Square Error from FTSM.restore Method

with images from other protocols, and so introduce the requirement for an extra inter-modality registration step.

References

1. D. C. Alexander. *An introduction to computational diffusion MRI: the diffusion tensor and beyond*. Chapter in 'Visualization and image processing of tensor fields' editted by J.Weickert and H.Hagen, Springer, 2006.
2. P. Basser, J. Mattiello & D. LeBihan. "MR diffusion tensor spectroscopy and imaging." *Biophysical Journal* **66**, pp. 259–67, 1994.
3. L.-C. Chang, D. K. Jones & C. Pierpaoli. "Restore: Robust estimation of tensors by outlier rejection." *Magnetic Resonance in Medicine* **53**, pp. 1088–1095, 2005.
4. P. Jezzard, A. S. Barnett & C. Pierpaoli. "Characterisation of and correction for eddy current artefacts in echo planar diffusion imaging." *Magnetic Resonance in Medicine* **39**, pp. 801–812, 1998.
5. M. Jenkinson, P. Bannister, M. Brady et al. "Improved optimization for the robust and accurate linear registration and motion correction of brain images." *NeuroImage* **17**, pp. 825–841, 2002.
6. J. C. Haselgrove & J. R. Moore. "Correction for distortion of echo-planar images used to calculate the apparent diffusion coefficient." *Magn. Reson. Med.* **36**, pp. 960–964, 1996.
7. M. Bastin. "Correction of eddy current-induced artifacts in diffusion tensor imaging using iterative cross-correlation." *Magn Reson Imaging* **17**, pp. 1011–1024, 1999.
8. T. Netsch & A. van Muiswinkel. "Quantitative evaluation of image-based distortion correction in diffusion tensor imaging." *IEEE Tansactions on medical imaging* **23(7)**, pp. 789–798, July 2004.
9. K. Jansons & D. Alexander. "Persistent angular structure: new insights from diffusion magnetic resonance imaging data." *Inverse Problems* **19**, pp. 1031–1046, 2003.
10. V. J. Wedeen, P. Hagmann, W.-Y. I. Tseng et al. "Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging." *Magnetic Resonance in Medicine* **54(6)**, pp. 1377–1386, December 2005.

Consistent Spherical Parameterisation for Statistical Shape Modelling

Rhodri H Davies *, Carole J Twining and Chris J Taylor

Division of Imaging Science and Biomedical Engineering (ISBE),
Faculty of Medicine, University of Manchester, UK

Abstract. We have described previously a method of automatically constructing statistical models of shape. The method treats model-building as an optimisation problem by re-parameterising each shape so as to minimise the description length of the training set. The approach requires an explicit parameterisation of each shape, which is straightforward in 2D, but non-trivial in 3D. It is necessary to provide some parameterisation of the training set to initialise the optimisation. An inappropriate initial parameterisation can cause the optimisation to converge at a slower rate or stop it from converging to a satisfactory solution. In this paper we describe a method of producing a consistent parameterisation for a given set of surfaces. The consistent parameterisations were used to initialise the model-building algorithm and produced results that were significantly better than alternative approaches.

1 Introduction

Statistical shape models (SSMs) have proven to be a powerful basis for image segmentation and shape analysis. Model-building involves establishing, from a training set, the pattern of 'legal' shape variation for a given class of object. Statistical analysis is used to provide an efficient parameterisation of this variability. The resulting models have good specificity (they can only represent valid instances of the class of object) and generalisability (they can represent any instance of the class of modelled object).

A key issue in constructing a shape model is establishing a dense correspondence between shape boundaries/surfaces over a reasonably large set of training images. It is important to establish the 'correct' correspondences, otherwise an inefficient parameterisation of shape can result, leading to non-specific models and non-general models. We have shown previously [1–3] how the correspondence problem can be solved by treating it as an integral part of the shape modelling process. The approach involves optimising the description length of the training set by explicitly manipulating the parameterisation of each training shape, leading to a training set with minimum description length (MDL).

The MDL method works robustly in 2D [1], but we have found that it can be sensitive to initialisation in 3D. The problem arises since the model-building algorithm requires an explicit parameterisation of each shape, which is straightforward in 2D, but non-trivial in 3D. If the initial parameterisations are inconsistent across the training set, then the initial correspondence is poor, which can slow down convergence of the MDL method or stop it from converging at all.

There are many methods of surface parameterisation; a review can be found in [4]. Each method produces a parameterisation with different properties such as a harmonic mapping or a conformal mapping, but none of the methods produce a parameterisation with the properties that we desire, namely one that is consistent between different instances of the same class of object.

In this paper, we treat surface parameterisation as an explicit optimisation problem. This approach to parameterisation allows us to construct consistent parameterisations that are guaranteed to have pre-determined desirable properties. The approach is simple yet effective and, as reported in section 5, leads to improved shape models.

2 Statistical Shape Models

A statistical shape model is built from a training set of example shapes $\{S_i : i = 1, \ldots n_s\}$, aligned to a common frame of reference. Each shape, S_i, can – without loss of generality – be represented by a set of n *corresponding* points regularly sampled on the shape. The concatenated coordinates of these points form the representation of the shape, an n_p dimensional shape vector $\mathbf{x_i}$. Using Principal Component analysis, each shape vector can then be expressed using a linear model of the form:

$$\mathbf{x_i} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b_i} = \bar{\mathbf{x}} + \sum_{\mathbf{m}} \mathbf{p^m} \mathbf{b_i^m}, \tag{1}$$

* firstname.lastname@manchester.ac.uk

where $\bar{\mathbf{x}}$ is the mean shape vector, $\mathbf{P} = \{\mathbf{p}^m\}$ are the eigenvectors of the covariance matrix that describe a set of orthogonal modes of shape variation and $\mathbf{b} = \{b^m\}$ are shape parameters.

The ability of the model to represent the shape variation depends crucially on the set of corresponding points chosen. The approach we will follow is to parameterise each shape. Corresponding points are then points on different shapes with the same parameter values.

3 Automatic Model-building by Direct Optimisation of Description Length

We have shown previously that the correspondence problem can be solved by treating it as an integral part of the shape learning process both in 2D [1, 2] and in 3D [3]. The basic idea is to choose the correspondences that build the 'best' model, treating model building as an optimisation task. This requires a framework involving:

• a method of *manipulating correspondences*,

• an *objective function* to assess the 'quality' of the model built from a given set of correspondences,

• a method of *optimising* the objective function with respect to the set of correspondences.

As explained in the previous section, we parameterise each shape. For a 3D surface with spherical topology, this involves mapping the surface of the shape to the surface of a sphere, with no folds or tears. The parameters are then just the usual spherical polar coordinates, and corresponding points across the set of shapes are points with the same values of the parameters. So, we now see how we can change the correspondence of a single shape S_i; we just need to re-parameterise that shape. That is, we need to apply a smooth, continuous transformation to the entire surface of the parameter sphere for just that shape. In what follows, we generate such transformations using a composition of clamped plate spline (CPS) transformations [5]. These are transformations which apply a homeomorphic mapping (no folds or tears) which deforms a cap-shaped region of the sphere. The parameters of a single transformation are the position of this region ($\mathbf{P}$), the size of this region (ω), and the amount of deformation ($\mathbf{a}$).

The parameterisation, and hence the correspondence, of each shape can then be individually adjusted. The objective function for this re-parameterisation is one based on the information-theoretic concept of Minimum Description Length (MDL). It has been shown previously that this produces models with good generalisation ability, specificity and compactness. Full details can be found in [1].

Optimisation proceeds in an iterative manner. At each iteration, we add N_K CPS warps to the parameterisation of each shape. The widths $\{\omega\}$ and positions $\{\mathbf{P}\}$ are chosen stochastically and fixed during optimisation; the amounts of deformation $\{\mathbf{a}\}$ are used as the parameters of the optimisation. The position of the clamped plate spline is chosen from a uniform distribution over the sphere and the width is chosen from the positive half of a Gaussian distribution with zero mean and standard deviation σ_G. The convergence of the algorithm is relatively insensitive to the value of σ_G. A value of $\sigma_G = \frac{1}{2}$ was used in the experiments reported below. Optimisation is achieved by estimating the gradient of the objective function with respect to the set $\{a\}$ [6] and performing a line search along this gradient to find the optimal objective function value.

As in many optimisation tasks, the initialisation is important; in our case, this is the initial parameterisation of the set of shapes. It is important to choose a consistent set of initial parameterisations, otherwise convergence of the optimisation is slowed down. The method of selecting a suitable initial parameterisation is the main contribution of this paper.

4 Consistent Surface Parameterisation

We start from a set of training shapes $\{S_i\}$ with spherical topology . Each shape is a triangulated surface in 3D, with vertices $\{\mathbf{v_j}\}$ and triangles $\{\mathbf{t_k}\}$, where $\mathbf{v}(\mathbf{t_k})$ are the vertices of triangle k. A parameterisation of a surface then consists of assigning a parameter value $\mathbf{u_j} \in \mathbb{S}^\mathbf{2}$ to each vertex $\mathbf{v_j}$. Barycentric coordinates are used to interpolate this parameterisation in between vertices.

An initial discrete parameterisation of the training surfaces can be achieved using many methods [4]. These parameterisation methods do not, however, consider the training set as a whole, hence the parameterisations are likely to be inconsistent across the group. Figure 1 shows a comparison between a SPHARM parameterisation, produced by the

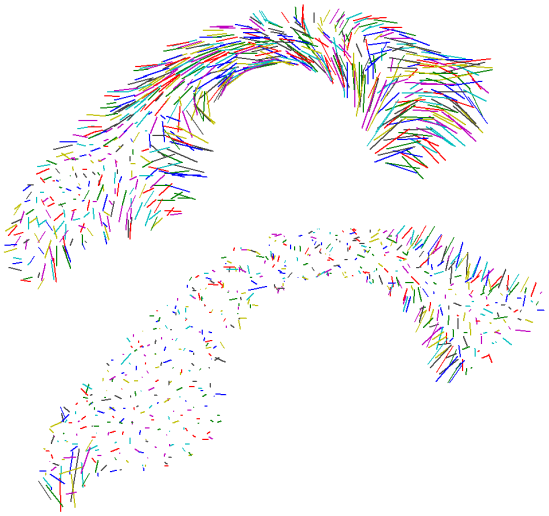


Figure 1. The correspondence (coloured lines are drawn between corresponding points) of a pair of brain ventricles. *Top*: SPHARM parameterisation; *Bottom*: groupwise consistent parameterisation.

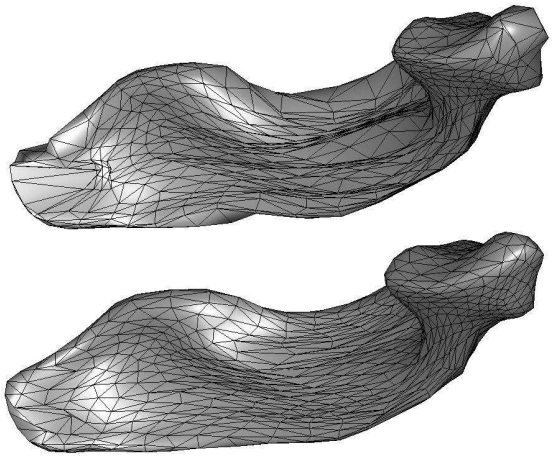


Figure 2. An example of the brain ventricle sampled according to its parameterisation before (*Top*) and after (*Bottom*) correcting for area distortion.

method of Brechbühler *et al.* [7], and a consistent parameterisation, produced by the method of Section 4.1.

Consistency of the parameterisation across the group is not the only consideration – parameterisation can also lead to (possibly severe) area distortion. This becomes a problem when model-building, since area distortion in the parameterisation leads to under- or over-sampling of the surface, giving a poor representation of the shape. This is because sampling is performed evenly on the parameter space (the sphere), rather than directly on the shapes. The benefit of correcting area distortion is shown in figure 2, which compares a surface sampled using two different parameterisations: a parameterisation with area distortion and a parameterisation that has been corrected using the method described below (Section 4.2).

4.1 Groupwise-consistent Parameterisation

We wish to manipulate the set of parameterisations so as to produce a good correspondence across the training set. A quick solution is to take each training surface in turn, identify a correspondence between it and the surface of a 'reference shape', and then manipulate the parameterisation to match this correspondence.

For the results reported below, the first training example was chosen as the reference shape, and its triangulated surface was decimated to give a new triangulation with 20% of the original vertices remaining.

Let $\{\mathbf{v}_i^A\}$ be the vertices of the decimated reference shapes and let $\{\mathbf{v}_j^B\}$ be the vertices of another of the training shapes. We assign a correspondence between each point, $\mathbf{v}_i^B$, and the closest point in $\{\mathbf{v}_j^B\}$. The result is an index, I, that maps the two sets such that $\mathbf{v}_{I(i)}^B$ is the closest point to $\mathbf{v}_i^A$.

We now wish to transform the parameterisation of shape B so that the parameter values $\{\mathbf{u}_{I(i)}^B\}$ best match the values of $\{\mathbf{u}_i^A\}$. This is achieved by manipulating $\{\mathbf{u}_{I(i)}^B\}$ so as to minimise a sum of squares objective function:

$$F = \sum_i |\mathbf{u}_i^A - \mathbf{u}_{I(i)}^B|^2. \tag{2}$$

We manipulate $\{\mathbf{u}_{I(i)}^B\}$ using a clamped plate spline-based representation of transformation, using the same algorithm as described in Section 2 above.

4.2 Minimising Distortion

Since the parameterisations are in correspondence, area distortion can be corrected by transforming the parameterisation on a single shape and then applying the same transformation to all members of the training set.

Area distortion is corrected by manipulating the parameterisation so as to minimise the following objective function:

$$F = \sum_{k}^{N_T} \left(\frac{area(\mathbf{v}(\mathbf{t}_k))}{\sum_j^{N_T} area(\mathbf{v}(\mathbf{t}_j))} - \frac{area(\mathbf{u}(\mathbf{t}_k))}{\sum_j^{N_T} area(\mathbf{u}(\mathbf{t}_j))} \right)^2, \tag{3}$$

where $area(\mathbf{v}(\mathbf{t}_j))$ is the area of the j^{th} triangle in $\mathbb{R}^3$ and $area(\mathbf{u}(\mathbf{t}_j))$ is the area of the same triangle in $\mathbb{S}^2$; N_T is the number of triangles. Manipulation of parameterisation and optimisation are as described in Section 2.

Note that the SPHARM method also produces parameterisations with no area distortion. In the results reported below, SPHARM was used as an initial estimate of parameterisation, hence there was very little distortion to remove. The distortion-minimising step is, however, likely to be critical if other methods are used as an initial estimate of parameterisation.

5 Results

MDL shape models of brain ventricles were constructed from 15 examples, using two different parameterisations: one obtained using the SPHARM method [7] and one using the Groupwise-consistent method of parameterisation proposed in this paper. To distinguish between the two, we will refer to them as 'SPHARM' and 'Groupwise-consistent' models respectively.

Qualitative results were produced by varying the first mode of variation by ± 2[standard deviations found over the training shapes]. A quantitative evaluation of the models was performed using two objective measures of model quality: generalisation ability and specificity. The derivation of each measure is given in [8]. Generalisation ability was measured by leave-one-out reconstruction, and specificity was assessed by generating a population of instances using the model and comparing them to the members of the training set. In both cases, lower values are preferable. The standard error of each measure can also be calculated, giving the significance of differences between models produced by different approaches.

Qualitative results in figure 3 show that the Groupwise-consistent model displays modes of shape variation that would be expected from the brain ventricle. The value of the MDL objective function is 80.3 for the SPHARM model and 68.4 for the Consistent Parameterisation model, suggesting that the Groupwise-consistent model performed better. The quantitative results in figure 4 confirm this by showing that the Groupwise-consistent model has significantly better specificity and generalisation ability than the SPHARM model.

6 Conclusions

We have described a simple yet effective method of achieving a consistent parameterisation of a set closed surfaces. The new method of parameterisation produces significantly better MDL shape models than alternative approaches to parameterisation.

We have also run the method described here on a set of proximal femurs and the resulting model was a vast improvement on the initial parameterisation. Future work will see us test the robustness of the method on a wider range of objects.

The MDL model-building algorithm would ideally be able to converge quickly and reliably from any given parameterisation of the training surfaces. The algorithm can, however, get trapped in local minima, which makes a consistent initial parameterisation essential. The focus of our current work is on investigating ways of making the model-building algorithm more robust to local minima.

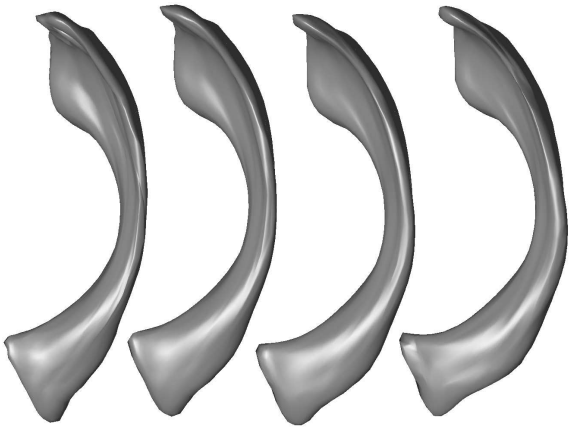


Figure 3. The first mode of variation of models built from the groupwise consistent parameterisation. Variation is plus and minus two standard deviations about the mean. The main variation is a change in curvature of the body and 'rotation' of the anterior horn.

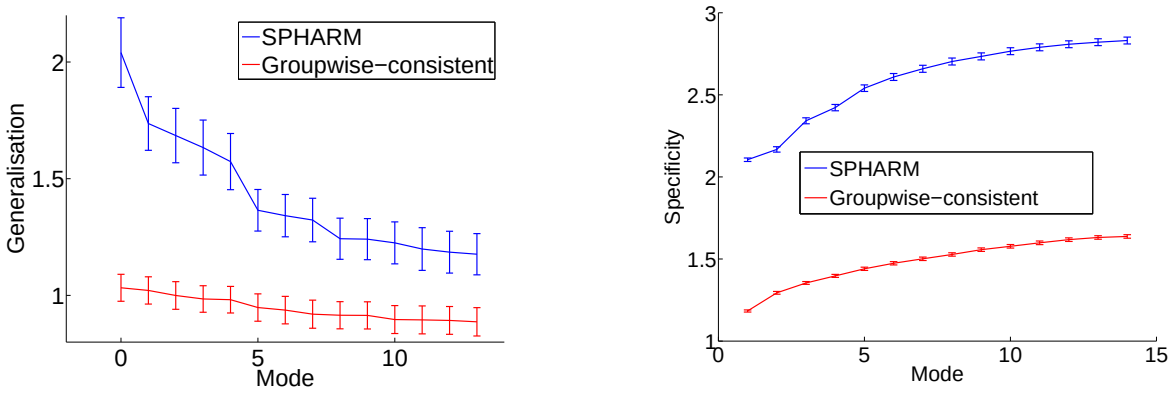


Figure 4. A quantitative comparison of the SPHARM and Groupwise-consistent models.

Acknowledgements

This work was supported by the MIAS IRC project; EPSRC grant number GR/N14248/01. Thanks to Martin Styner and Guido Gerig (UNC, Chapel Hill) and Daniel Weinberger and Douglas Jones (NIMH, Bethesda) for the Ventricle Data.

References

1. R. H. Davies, C. J. Twining, T. F. Cootes et al. "A minimum description length approach to statistical shape modelling." *IEEE Transactions on Medical Imaging* **21(5)**, pp. 525–537, May 2002.
2. R. H. Davies, C. J. Twining, P. D. Allen et al. "Constructing optimal 2d statistical shape models." In *Image and Vision Computing*, volume 21, pp. 1171–1182. 2003.
3. R. H. Davies, C. J. Twining, T. F. Cootes et al. "3d statistical shape models using direct optimisation of description length." In 7^{th} *European Conference on Computer Vision*, pp. 3–21. 2002.
4. M. S. Floater & K. Hormann. "Surface parameterization: a tutorial and survey." *Advances in Multiresolution for Geometric Modelling* pp. 157–186, 2004.
5. C. Twining, S. Marsland & C. Taylor. "Measuring geodesic distances on the space of bounded diffeomorphisms." In *bmvc 2002*, pp. 847–856. 2002.
6. T. Heimann, I. Wolf, T. G. Williams et al. "3d active shape models using gradient descent optimization of description length." In 19^{th} *Conference on Information Processing in Medical Imaging*, pp. 566–577. 2005.
7. C. Brechbühler, G. Gerig & O. Kühler. "Parametrization of closed surfaces for 3d shape description." *Computer Vision and Image Understanding* **61(2)**, pp. 154–170, 1995.
8. R. H. Davies. *Learning Shape : Optimal Models of Natural Variability*. Ph.D. thesis, University of Manchester, UK, 2002.

Probabilistic Segmentation of the Knee Joint from X-ray Images

Matthias Seise[1]*, Stephen J. McKenna[1], Ian W. Ricketts[1] and Carlos A. Wigderowitz[2]

[1]Division of Applied Computing, University of Dundee, Dundee, Scotland, DD1 4HN
[2]Orthopaedic and Trauma Surgery, Ninewells Hospital and Medical School, Dundee, Scotland, DD1 9SY

Abstract. A probabilistic method is proposed for segmentation of the knee joint. A likelihood function is formulated that explicitly models overlapping object appearance. Priors on global appearance and geometry (including shape) are learned from example images. Markov chain Monte Carlo methods are used to obtain samples from a posterior distribution over model parameters from which expectations can be estimated. The result is a probabilistic segmentation that quantifies uncertainty so that measurements such as joint space can be made with associated uncertainty. Joint space area and mean point-to-contour distance are used for evaluation.

1 Introduction

The aim of this paper is to outline a probabilistic, model-based segmentation method for the knee joint from x-ray images and to make explicit the uncertainty in the segmentation so obtained. The method explicitly handles the possible overlapping of femur and tibia and their appearance models. Such cases are not handled in methods based on active contours [1], active shape models or active appearance models [2], for example.

Segmentation of objects is often only an intermediate result. Consider for example medical image analysis tasks which involve measuring the size of anatomical structures. Most standard segmentation algorithms result in a single solution without any information as to the confidence in this solution. No information about uncertainty is propagated to the subsequent size estimation step. Furthermore, anatomical structures almost inevitably overlap. In medical applications, especially, it is desirable to have an indication of the certainty of a measurement and to cope with structures that overlap or are in close proximity. The performance of the proposed method is evaluated by applying it to the segmentation of the knee joint to enable the measurement of joint space, an important biomarker for the assessment of osteoarthritis [3].

2 Modelling Knee Radiographs

The general task of segmenting modelled objects from an image can be described in a Bayesian framework as that of inferring the conditional distribution $P(\mathcal{M}|\mathcal{I},\mathbb{I},\mathbb{S})$ which is the probability of the model parameters, $\mathcal{M}$, given a test image $\mathcal{I}$, a set of training images, $\mathbb{I}$, and their annotations, $\mathbb{S}$. Each object can be described by its geometry and appearance in the image and, more specifically, in terms of shape parameters, S, geometric (non-shape) parameters, G, global appearance parameters, A_g, and local appearance parameters, A_l. Using Bayes' rule and assuming $P(\mathcal{I})$ is fixed:

$$P(\mathcal{M}|\mathcal{I}) \propto p(\mathcal{I}|\mathcal{M})P(\mathcal{M}) = p(\mathcal{I}|G,S,A_g,A_l)P(G)P(S)P(A_g)P(A_l) \tag{1}$$

Assuming pixel-wise conditional independence, the likelihood factors as

$$p(\mathcal{I}|G,S,A_g,A_l) = \prod_{n=1}^{N}\prod_{x\in\Omega_n} p_n(\mathcal{I}(x)|G,S,A_g,A_l) \tag{2}$$

where the likelihood is modelled as not depending on pixels outside a finite image support $\Omega = \bigcup_{n=1}^{N}\Omega_n \subset \mathbb{R}^2$ consisting of N non-intersecting regions.

Point distribution models were used to model shape deformations in femoral and tibial contours [2]. Specifically, each normalised femoral shape z was represented using shape parameters S_f by $z \approx \overline{z} + \Phi(\Lambda^{\frac{1}{2}} \bullet S_f)$ with $\overline{z}$ being the mean shape, Φ the eigenvectors and $\Lambda^{\frac{1}{2}}$ the square roots of the corresponding eigenvalues of the covariance matrix, all learned from training data. $(\Lambda^{\frac{1}{2}} \bullet S_f)$ denotes the componentwise product. Shape parameters, S_t, for tibia were defined analogously. Further geometric parameters were global parameters $G_g = \{t_x, t_y, \theta, s_f\}$ where (t_x, t_y) is the midpoint of the line segment connecting the centres of mass of the femur and tibia, θ is the rotation of this line segment, and s_f is scale, and local parameters $G_l = \{d, s_t, \theta_f, \theta_t\}$ where d is the separation of the bones, s_t is the scale of the tibia relative to the femur, and θ_f and θ_t are local rotations of the femur and tibia (see Figure 1)

X-ray images can vary a lot in terms of brightness and contrast as well as in the imaged region of anatomy and surrounding area. Rather than an appearance model that accounts for the entire image it is sensible to consider only a limited region around the bone contours. Specifically, appearance models were bands of width $2h$ centred on the contour. These can be described by warp functions $\omega : x \in \mathbb{R}^2 \mapsto (\alpha, \delta) \in [0,1] \times [-h, h]$ from subsets of image coordinates to rectangular regions, the texture patches. The value of α corresponds to the normalised arclength of the shape and δ to the distance normal to the shape. It is important to note that ω is dependent on the shape and position of the object. Global appearance parameters, brightness β and contrast γ, are included in the model parameters as $A_g = \{\beta, \gamma\}$ and are thus automatically inferred by the segmentation algorithm.

*Corresponding author: M. Seise. email: mseise@computing.dundee.ac.uk

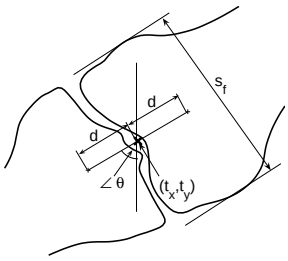


Figure 1. Geometric model parameters

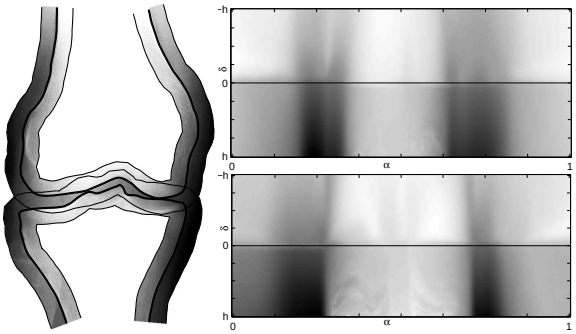


Figure 2. Local appearance. Left: masks. Right: associated texture patches (means only)

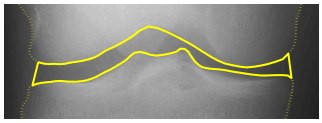


Figure 3. Definition of joint space area.

2.1 Single Object Appearance

A warp ω maps intensity values of an image region Ω onto the rectangular texture patch $T : [0,1] \times [-h,h] \rightarrow \mathbb{R}$ by $T(\omega(\Omega)) = \mathcal{I}(\Omega)$. Making the assumption that pixel values on a texture patch are independently Gaussian distributed, the local appearance is modelled by pixelwise means and variances, $A_l = \{\boldsymbol{\mu}_T, \boldsymbol{\sigma}_T^2\}$ which are estimated from training data using maximum likelihood[1]. The likelihood of the parameters is modelled using a product of Gaussians:

$$p(\mathcal{I}|G,S,A_g,A_l) = \left(\prod_{x \in \Omega} p\left(\frac{\mathcal{I}(x) - \beta}{\gamma} \middle| \boldsymbol{\mu}_T(\omega(x)), \boldsymbol{\sigma}_T^2(\omega(x)) \right) \right)^{\frac{W}{|\Omega|}} \tag{3}$$

where geometry and shape are accounted for by the warp ω and to avoid the need to always use the same number of observations, the function is normalised with W chosen according to a notional mask size. This weighting enables the importance of the prior to be varied. This was used subsequently in the MCMC algorithm to achieve sensible acceptance rates.

2.2 Multiple Object Appearance

When segmenting multiple objects, appearance overlap should be taken into account. Each object's appearance mask divides the image into three image regions. The region corresponding to the mask divides into regions hypothesised as interior and exterior to the object. The third image region is the background not corresponding to the mask. These regions are named $\Omega_f^{in}, \Omega_f^{out}, \Omega_f^{bgd}$ for the femur and $\Omega_t^{in}, \Omega_t^{out}, \Omega_t^{bgd}$ for the tibia. Considering object overlaps, there are nine possible region types: $\Omega_f^{in} \cap \Omega_t^{in}, \Omega_f^{in} \cap \Omega_t^{out}, \ldots, \Omega_f^{bgd} \cap \Omega_t^{bgd}$. For each of these possibilities, likelihood functions must be chosen. Those used here are motivated by x-ray image formation.

2.2.1 X-ray image formation

Photons emitted from an x-ray source and interacting with tissue are absorbed, scattered or transmitted. Noise due to scattering is largely filtered out in modern x-ray radioscopic equipment. The probability P that a photon is absorbed follows Beer's Law, so $P = e^{-m}$ where $m = \int_L \xi(x)dx$, with ξ being the opacity and L the path of the photon through the object. In the simplest case $m = \xi \cdot l$ with l being the thickness of the object and ξ the mean opacity. The intensity I of a point in an x-ray image created by a ray that has passed through an object is then approximated as $I = I_0 e^{-m}$ where I_0 is the intensity of a pixel created by an undisturbed ray. (The film, equipment and digitisation process also influence the image but are not modelled here.) Consider the case of two objects that are partially overlapping. An x-ray passing only through the first object creates an intensity $I_1 = I_0 e^{-m_1}$ and one passing only through the second object creates $I_2 = I_0 e^{-m_2}$. The intensity $I_{1,2}$ where the objects' images overlap is then

$$I_{1,2} = I_0 e^{-(m_1+m_2)} = \frac{I_1 I_2}{I_0}$$

Without loss of generality let the maximum intensity in an image $\mathcal{I} : \mathbb{R}^2 \rightarrow [0,1]$ be 1. It can be assumed that at least one pixel in an image is created by an undisturbed ray. Therefore, $I_0 = 1$ and intensities in regions of overlap are approximately multiplicative.

[1]For N texture patches, $\{T_1, \ldots, T_N\}$, $\boldsymbol{\mu}_T$ is defined as the "pixelwise" mean $\boldsymbol{\mu}_T(\alpha,\delta) = \frac{1}{N} \sum_{n=1}^{N} T_n(\alpha,\delta)$ and $\boldsymbol{\sigma}_T^2$ as the "pixelwise" variance $\boldsymbol{\sigma}_T^2(\alpha,\delta) = \frac{1}{N-1} \sum_{n=1}^{N} (T_n(\alpha,\delta) - \boldsymbol{\mu}_T(\alpha,\delta))^2$.

$x \in$	Likelihood factor	abbr.[a]
Ω_1	$p\left(\frac{\mathcal{I}(x)-\beta}{\gamma}\middle\vert \boldsymbol{\mu}_{T_f}(\omega(x)), \boldsymbol{\sigma}^2_{T_f}(\omega(x))\right)$	$p_f(x)$
Ω_2	$p\left(\frac{\mathcal{I}(x)-\beta}{\gamma}\middle\vert \boldsymbol{\mu}_{T_t}(\omega(x)), \boldsymbol{\sigma}^2_{T_t}(\omega(x))\right)$	$p_t(x)$
Ω_3	$\frac{1}{2}p\left(\frac{\mathcal{I}(x)-\beta}{\gamma}\middle\vert \boldsymbol{\mu}_{T_f}(\omega(x)), \boldsymbol{\sigma}^2_{T_f}(\omega(x))\right) + \frac{1}{2}p\left(\frac{\mathcal{I}(x)-\beta}{\gamma}\middle\vert \boldsymbol{\mu}_{T_t}(\omega(x)), \boldsymbol{\sigma}^2_{T_t}(\omega(x))\right)$	$p_{f+t}(x)$
Ω_4	$\mathcal{F}(\boldsymbol{\mu}_{T_f}(\omega(x)), \boldsymbol{\sigma}^2_{T_f}(\omega(x)), \boldsymbol{\mu}_{T_t}(\omega(x)), \boldsymbol{\sigma}^2_{T_t}(\omega(x)))$	$p_{\mathcal{F}}(x)$

[a] Abbreviation to simplify further notation.

Table 1. Likelihood calculation with overlaps ($\mathcal{F}$ is the normal product distribution)

2.2.2 Likelihood functions

The mapping from an observation's region type to its associated likelihood function is given in Table 1. Since the opacity of soft tissue is smaller than that of bone, the likelihood for pixels lying in an overlap of soft tissue and bone is approximated as being the same as if only bone were present. In regions of overlap of the exterior parts of two object masks, a Gaussian mixture (abbreviated as $p_{f+t}(x)$) is used since the observations are modelled by both masks. In regions of overlap of the interior parts of two object masks, the intensities are approximately multiplicative, so the *normal product distribution* $\mathcal{F}$ is used. Usually its density *cannot* be calculated analytically but can be approximated by $\mathcal{N}(\boldsymbol{\mu}_{T_f}(\omega(x))\boldsymbol{\sigma}^2_{T_f}(\omega(x)), \boldsymbol{\sigma}^2_{T_f}(\omega(x))\boldsymbol{\sigma}^2_{T_t}(\omega(x)) + \boldsymbol{\mu}_{T_f}(\omega(x))\boldsymbol{\sigma}^2_{T_t}(\omega(x)) + \boldsymbol{\sigma}^2_{T_f}(\omega(x))\boldsymbol{\sigma}^2_{T_t}(\omega(x)))$ [4]. The final likelihood is then

$$p(\mathcal{I}|G, S, A_g, A_l) = \left(\prod_{x\in\Omega_1} p_f(x) \cdot \prod_{x\in\Omega_2} p_t(x) \cdot \prod_{x\in\Omega_3} p_{f+t}(x) \cdot \prod_{x\in\Omega_4} p_{\mathcal{F}}(x)\right)^{\frac{W}{V}} \tag{4}$$

with

$$\Omega_1 = (\Omega_f^{in} \cap (\Omega_t^{out} \cup \Omega_t^{bgd})) \cup (\Omega_f^{out} \cap \Omega_t^{bgd}) \qquad \Omega_2 = (\Omega_t^{in} \cap (\Omega_f^{out} \cup \Omega_f^{bgd})) \cup (\Omega_t^{out} \cap \Omega_f^{bgd})$$

$$\Omega_3 = \Omega_f^{out} \cap \Omega_t^{out} \qquad \Omega_4 = \Omega_f^{in} \cap \Omega_t^{in} \qquad V = \left|\bigcup_{i=1}^{4} \Omega_i\right|$$

Learning the likelihood model Appearance learning constructs the texture patches $\{T_1, \ldots, T_n\}$ and calculates the pixelwise means $\{\boldsymbol{\mu}_T\}$ and variances $\{\boldsymbol{\sigma}^2_T\}$. For our application, only regions not overlapping other bones were used to estimate them. Missing data were then estimated by interpolating from neighbouring pixels. Estimated means are shown in Figure 2.

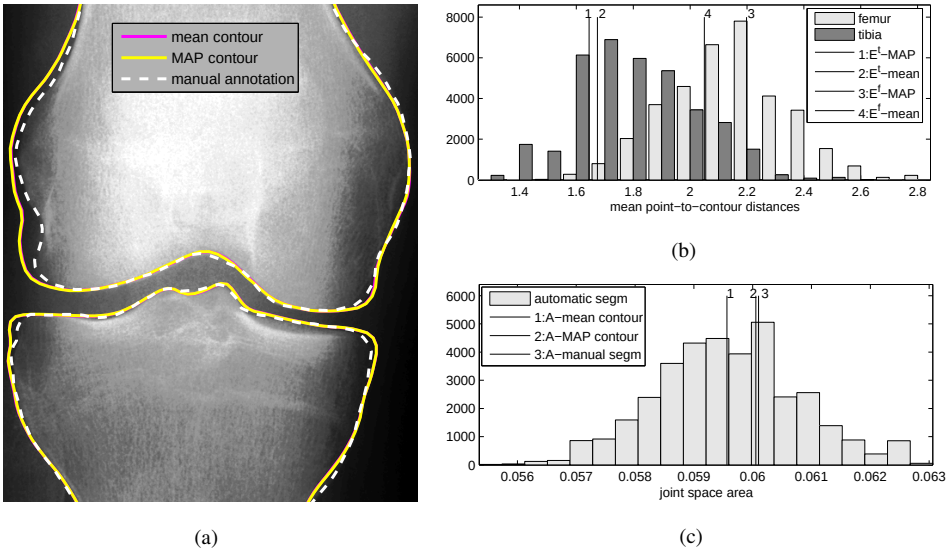
2.3 Prior Distributions

Priors $P(G), P(S), P(A_g), P(A_l)$ were chosen to reflect beliefs about the behaviour of the variables. Gaussian priors were used for brightness β, contrast γ, rotations $\theta, \theta_f, \theta_t$ and local appearance A_l with the corresponding means and variances estimated from training data using maximum likelihood. Translation (t_x, t_y) had a uniform prior over a rectangular region of the image and scale priors for s_f and s_t were log-Gaussian. The specifying parameters of the latter distributions were also estimated from training data.

The learning of the shape model priors was similar to the ASM shape model building step [2] and was performed independently for femur and tibia. Femur shapes $\mathbb{S}_f = \{s_f^1, \ldots, s_f^N\}$ were aligned and normalised using Procrustes analysis. The mean $\overline{S}_f$ and the covariance matrix Σ_f were calculated. PCA was applied to Σ_f decomposing it into $\Sigma_f = \Lambda_f \cdot \Phi_f \cdot \Lambda_f^{-1}$, the eigenvalue vector Λ_f and the eigenvectors Φ_f. With the standard assumption of a Gaussian shape space it is implied that the prior of the shape parameters S_f is the multivariate standard normal distribution. Tibia shapes $\mathbb{S}_t$ were processed analogously.

2.4 Sampling the Posterior with Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are used to generate samples from a target distribution and subsequently to estimate expectations of functions under this distribution (see Andrieu *et al.* [5] for a good introduction). To sample from the posterior, Equation (1), we used a mixture of Metropolis-Hastings random walk MCMC and a variant of hybrid MCMC. Geometric parameters, G, were updated componentwise and shape parameters blockwise using Gaussian random walk. Blocks were chosen such that each block accounted for approximately the same variance. Global appearance parameters $A_g = \{\beta, \gamma\}$ were updated together using the Langevin method since partial derivatives with respect to brightness $\frac{\partial P(\mathcal{I}|\mathcal{M})}{\partial \beta}$ and contrast $\frac{\partial P(\mathcal{I}|\mathcal{M})}{\partial \gamma}$ are easily and efficiently calculated.



(a) (b) (c)

Figure 4. (a) Example segmentation result with corresponding histograms of (b) mean point-to-contour distances and (c) joint space areas for $40,\ 000$ MCMC samples.

3 Experiments

Methods were evaluated on a data set of 30 standard clinical x-rays (see also [3]). Images of left knees were mirrored so that they appeared as right knees. All images were manually annotated and leave-one-out validation was used. Shapes were brought into correspondence using the MDL approach [6] with curvature [7]. After applying PCA, the 16 major modes were left unchanged. Rather than discarding the minor modes, their variances were set to the mean value of the remaining variances to enable search outside the lower-dimensional subspace.

A coarse-to-fine algorithm was used for speed since likelihood calculation based on this model was computationally expensive. A two-step approach was used to first find the main modes of the posterior at low resolution and then explore them at full resolution.

1st step: coarse resolution. Shape was fixed to the mean shape while the other geometric and the global appearance parameters were updated. The Markov chain was initialised randomly as a sample from the prior distribution. The standard deviations of the proposal functions were set equal to the standard deviations of the associated parameters learned from the training set. The weight W was set so that likelihood and prior were of comparable magnitude so as to encourage exploration of the posterior. (Typical performance: 20% acceptance rate, 8 minutes per 1000 samples on Centrino 1.9 GHz, 1GB RAM, implementation in Matlab.)

2nd step: fine resolution. The chain was initialised at the MAP solution obtained from step 1. The weight W was increased to encourage exploration of local modes. All model parameters were updated, geometric and global appearance parameters with decreased variance. Shape parameters were updated in blocks with proposal standard deviations set to half the standard deviations of the respective parameters. Further improvements in speed were achieved by sampling pixels in "uninteresting" regions of the mask at a lower density. The application of measuring the joint space demands higher accuracy along the femoral condyles and tibial plateaux than elsewhere. Therefore, the likelihood was calculated for every pixel around the joint space and only at quarter resolution in the other areas. Step 2 resulted in a representation of the posterior distribution (with the first 10% of samples discarded as burn-in). (Typical performance: 8% acceptance rate, 30 minutes per 1000 samples on Centrino 1.9 GHz, 1GB RAM, implementation in Matlab.)

Figure 4(a) shows an example test image with the mean and MAP estimates overlaid along with a manual annotation. The mean and MAP contours cannot be readily distinguished and this was also true of most other test images. Inferred contours can be compared to a manually annotated contour based on *mean point-to-contour distance* defined as the average Euclidean distance from the landmark positions to the annotated contour. Mean point-to-contour distances along the femoral condyles (E^f) and along the tibial plateaux (E^t) of the image in Figure 4(a) are shown in Figure 4(b). The values of E^f and E^t are also indicated for the mean and MAP contours computed from the Monte Carlo samples. Further evaluation is based on the joint space area. In the absence of a generally accepted definition of joint space area we used the area between the tibial plateaux and femoral condyles as shown in Figure 3, normalised by the overall scale of the bone. Figure 4(c) shows the calculated areas as well as those of the mean contour, the MAP contour and the manual annotation for the image in Figure 4(a).

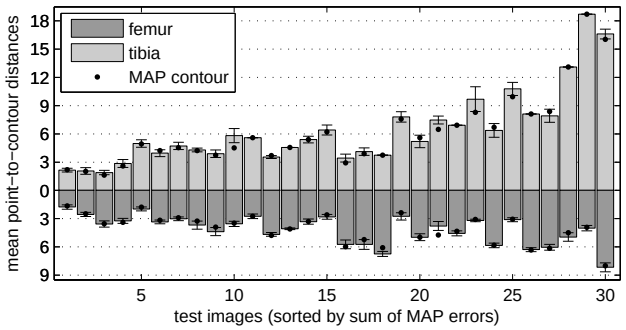


Figure 5. Statistics summarising mean point-to-contour distances for the test images (see text).

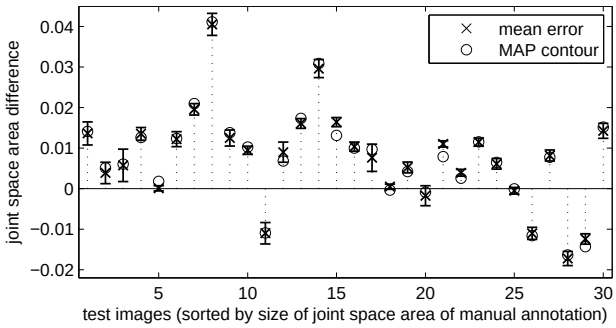


Figure 6. Statistics summarising difference in joint space area measurements (see text).

The values of E^f and E^t obtained with the thirty test images are summarised in Figure 5. Results for the femur are shown below the baseline and for the tibia above. Each histogram bar indicates the Monte Carlo estimate of the expectation of the mean point-to-contour distance under the posterior distribution. Error bars indicate standard deviations and MAP contour errors are also indicated. Analogously, the differences between the the joint space area of the manual annotation and of the Monte Carlo samples and the MAP contour are summarised in Figure 6. It shows that the automatic method tends to give larger areas than the human annotator.

4 Discussion and Conclusions

A probabilistic framework for segmentation of the knee joint was described in which appearance overlaps were explicitly modelled. The output was not a single contour but Monte Carlo samples from a posterior distribution over contours. The approach is broadly applicable since overlapping objects are almost inevitable in medical applications (e.g. [8]). Furthermore, since segmentation is often only an intermediate step it is important to quantify and propagate uncertainty to higher level processes. An important example is the estimation of areas and volumes in medical images. The method was applied to segmentation of femoral condyles and tibial plateaux and the joint space area was computed as a possible proxy for cartilage volume. It achieved reasonable results based on a relatively small and challenging data set. The accuracy of the joint space area measurement has to be confirmed in further studies. The current implementation would seem to have a tendency to overestimate joint space. However, it must be noted that, so far, neither errors in the manual annotation nor flaws in the definition of the joint space area were considered. Future work could extend the approach by explicitly modelling uncertainty in the manual annotations upon which learning is performed. Other opportunities include the use of data-driven MCMC schemes to ensure full exploration of the multi-modal posterior [9] and enhancement of the likelihood model.

Acknowledgments

M. Seise was partially funded by EPSRC. H. H. Thodberg made available code for the MDL method. Dr B. Oliver provided knee images.

References

1. M. Kass, A. Witkin & D. Terzopoulos. “Snakes: Active contour models.” *IJCV* **1(4)**, pp. 321–331, 1988.
2. T. F. Cootes & C. J. Taylor. “Statistical models of appearance for computer vision.” Tech. rep., University of Manchester, 2004.
3. M. Seise, S. J. McKenna, I. W. Ricketts et al. “Double contour active shape models.” In *BMVC*, pp. 159 – 168. 2005.
4. R. Ware & F. Lad. “Approximating the distribution for sums of products of normal variables.” Tech. rep., University of Canterbury, 2003.
5. C. Andrieu, N. de Freitas, A. Doucet et al. “An introduction to MCMC for machine learning.” *Machine Learning* **50**, pp. 5–43, Jan-Feb 2003.
6. R. H. Davies, C. Twining, T. F. Cootes et al. “A minimum description length approach to statistical shape modelling.” *IEEE Trans. Med. Imag.* **21(5)**, pp. 525 – 537, 2002.
7. H. H. Thodberg & H. Olafsdottir. “Adding curvature to minimum description length shape models.” In *BMVC*, pp. 251 – 260. 2003.
8. P. P. Smyth, C. J. Taylor & J. E. Adams. “Vertebral shape: automatic measurement with active shape models.” *Radiology* **211(2)**, pp. 571 – 578, 1999.
9. C. Liu, H.-Y. Shum & C. Zhang. “Hierarchical shape modeling for automatic face localization.” In *ECCV*, pp. 687–703. Springer-Verlag, London, UK, 2002.

Anatomically Corresponded Knee Cartilage Thickness Analysis from MRI

Tomos G. Williams[a*], Andrew P. Holmes[b], John C. Waterton[b],
Rose A. Maciewicz[b], Anthony P. Nash[b] and Chris J. Taylor[a]

[a]Imaging Science and Biomedical Engineering, University of Manchester, Manchester, U.K.
[b]AstraZeneca, Alderley Park, Macclesfield, Cheshire, U.K.

Abstract. Cartilage loss due to Osteoarthritis (OA) is small and localised, making it difficult to detect and measure. We present a method which defines an anatomically consistent frame of reference across a population by constructing statistical shape models of the knee bones. Bone and cartilage were manually segmented from Magnetic Resonance Images (MRI) of volunteers' knees. Dense correspondences were defined across all subjects by constructing Minimum Description Length (MDL) statistical shape models of the bones. Regions of interest were manually delineated on the mean bone shapes provided by the models, and propagated to each individual in an anatomically consistent manner, using the model-based correspondences. This approach results in precise measurements that can be used to detect small localised changes in cartilage thickness.

1 Introduction

Osteoarthritis (OA) is a widespread disease that has a degenerative effect on cartilage in the knee and other articulating joints [1]. Accurate measurement of the effects of OA on cartilage would increase understanding of the disease process and facilitate drug development through improved treatment monitoring. This is not straight-forward since the loss of cartilage is slow and restricted to particular sub-regions of the joint [2].

Magnetic Resonance Imaging (MRI), followed by advanced image analysis has proved effective in obtaining detailed cartilage thickness maps for individual knees [3]. Since cartilage loss tends to be localised, it is necessary to establish a fixed frame of reference, if disease progression is to be detected. In line with recent recommendations [4], we chose to use the underlying bone to provide a consistent frame of reference for each cartilage sheet, since it remains relatively stable during OA progression, whereas the cartilage itself is subject to change in both shape and topology.

Cohort studies are essential to understanding population effects, but require thickness maps for different individuals to be transformed into a common frame of reference. This requires the underlying bones to be brought into correspondence. This has been attempted previously using surface registration [5] or coordinate frame registration [6], but these methods do not account for population variation in joint shape, and thus fail to provide an anatomically meaningful correspondence.

In this paper we find a dense correspondence between the knee bones of different individuals, by constructing Minimum Description Length (MDL) statistical shape models of the bones. The method deals, explicitly, with shape variation across the population, and establishes anatomically meaningful correspondences. This allows regions of interest to be defined and propagated to individual bones in an anatomically consistent manner. Regional measurements of cartilage thickness are shown to be more reproducible than whole compartment measurements and, therefore, more likely to allow the effects of osteoarthritis on articular cartilage to be detected in cohort studies.

2 Method

2.1 Data Acquisition and Segmentation

Twenty healthy female volunteers (aged range 24-58) were imaged using 1.0 or 1.5T MRI with different vendors' instruments at three centres in the UK: Bristol-Siemens, Liverpool-GE, and Manchester-Philips. During each visit, two MR sequences were used: a fat-suppressed T1-weighted 3D gradient echo sequence for visualisation of the hyaline cartilage, and a T2-weighted sequence to define the end-osteal bone surface, both with $0.625 \times 0.625 \times 1.6$mm resolution. One dataset was excluded due to the use of an incorrect pulse sequence. Further details of the study are presented in [7]. Cartilage was segmented from the T1 weighted images using a region-growing algorithm implemented in TOSCA software (IBM, Winchester, U.K.). Two trained non-experts, under the supervision of an experienced segmentor, performed slice-by-slice segmentation of each cartilage image, two times each, resulting in a total of 76 cartilage

*tomos.williams@manchester.ac.uk

segmentations (19 patients $\times$ 2 segmentors $\times$ 2 repeat segmentations). Bones were segmented by a third non-expert, trained by a radiologist to manually trace the inner surface of the cortical bone in each slice of the T2 weighted images using EndPoint software (Imorphics, Manchester, UK).

2.2 Corresponding Cartilage Thickness Maps

To build detailed cartilage thickness maps, anatomically corresponded across a cohort, a common frame of reference was defined by establishing a dense correspondence between the bone surfaces of different subjects. The cartilage thickness was measured at the corresponding points. Here we present an overview of the method; a detailed description is presented in [8].

To enable 3D thickness measurements from any point on the bone surfaces, closed surface representations of the cartilage and bone were constructed from the slice-by-slice segmentations. Chemical shift and movement artifacts were corrected by rigid surface registration of the bones relative to the inner cartilage surface.

We have shown previously how a dense set of anatomically corresponding landmarks can be obtained by manipulating the positions of candidate points to give a Minimum Description Length (MDL) statistical shape model that optimally describes the set of shapes [9]. This provides a mean shape, a set of modes of shape variation, and a set of dense anatomical correspondences. MDL models were constructed for the femur, tibia and patella bones. Linear interpolation of the model correspondences was performed, as required, to ensure sufficiently dense coverage of measurement points and surface triangulation was performed. Measurement point density was measured as the mean separation of adjacent points averaged over all points and all shape examples.

Cartilage thickness was measured at each of the measurement points for each example. A 3D surface normal to the bone was defined at each point as the mean of the connecting facets' normals, weighted by the angle subtended at the point. Thickness was measured as the distance between the points of intersection of the bone normals with the inner and outer cartilage surfaces. Three criteria were used to test the validity of individual thickness measurement. First, the angle between the bone normal and cartilage surface normals at the points of intersection had to be less than 2π, discarding measurements from bone normal vectors which intersected the cartilage surfaces from above. Second, the distance from the bone to the inner cartilage surface had to be within the range [-3,+4mm]; this excluded measurements from cartilage not directly above the bone surface but included measurements where segmentation and/or surface registration errors resulted in the cartilage overlapping the bone. Finally, thickness measurements were restricted to the range [0,25mm] to eliminate extreme thickness measurements that occur when the bone normal is almost parallel to the cartilage surface.

2.3 Functional Regions

Changes in cartilage thickness in OA tend to be spatially heterogeneous, thus others have proposed regional analysis of cartilage morphology [10]. Current methods require the sub-compartments to be manually identified for each joint during image segmentation [4]. Here we exploit the dense set of correspondences between bones to automate the identification of anatomically consistent sub-compartments.

Sub-regions of the femoral and tibial articular surfaces were defined on the mean bone shapes according to the bearing surfaces of the different joint compartments. The tibia was divided into its lateral (LT) and medial (MT) aspects. The femur was divided into Trochlearfemur (TrF), the surface in contact with the patella, and Lateral Femorotibial (LF) and Medial Femorotibial (MF) sub-regions, the surfaces in contact with their respective tibial compartments. The dividing line between the trochlear and femorotibial sub-compartments was drawn from the inferior viewpoint on the mean bone shape as a straight line, perpendicular to the medial and tangential to the posterior aspect of the trochlear groove. The dividing boundaries were drawn along the facet edges connecting the measurement points, and could thus be propagated from the mean bone shape to any individual bone shape, automatically defining anatomically corresponding functional regions.

2.4 Subchondral and Trimmed Bone Regions

In characterising cartilage quantity, changes to both the thickness and extent of the cartilage sheet are of interest. Computing the mean cartilage thickness over the currently covered bone area does not account for denuded areas where previously existing cartilage has disappeared. It is therefore necessary to define a consistent region across all individuals over which to compute mean cartilage thickness. The maximal coverage region, denoted tAB, defined as the connected set of measurement points which exhibited cartilage coverage in any individual, was chosen as the

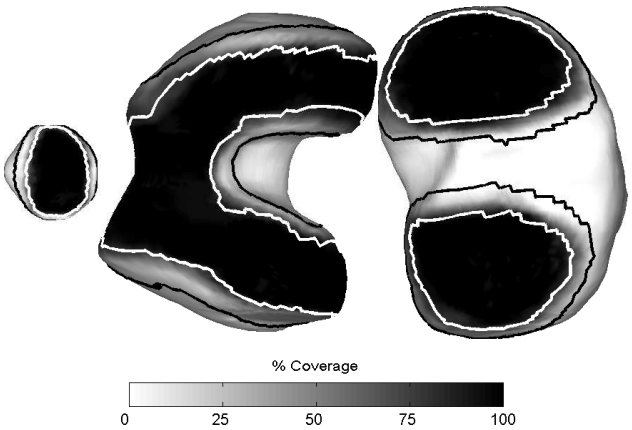


Figure 1. Aggregate cartilage coverage map with the subchondral and trimmed boundaries.

anatomically consistent region. This is an overestimation of the expected cartilage coverage region for any individual.

The edges of the cartilage sheets are least involved in OA disease and are difficult to segment, particularly, given the partial volume effects where the cartilage sheet curves into the image plane. To limit measurement to the central load-bearing sub-regions of each cartilage sheet, trimming boundaries were defined on the mean bone shapes. The trimmed region for each cartilage compartment, denoted ${}^{Trm}AB$, was drawn manually as a closed polygon along the facet edges of the mean bone surface, such that it tightly encompassed the set of points which exhibited cartilage coverage in at least 90% of the segmentations from the healthy female volunteers.

Since the subchondral and trimmed regions were defined along the facet boundaries they could be propagated to all individuals' bone surfaces using the model based correspondences.

2.5 Measuring Mean Cartilage Thickness

The most basic quantitative measure of cartilage is volume, calculated directly from the cartilage segmentations. Volume of cartilage, denoted VC, was computed for the major compartments from the closed cartilage surfaces using Gauss' theorem [11]. To normalise for variation in joint size, cartilage volume can be divided by the subchondral bone area to provide a measure, denoted VC/tAB, with units of millimetres [12].

Normalised cartilage volume measures have been used widely to measure cartilage morphology, but they require explicit delineation of boundaries on each individual if measurements are to be restricted to sub-regions of the joint [2]. In our framework it is possible to automate the process. We calculated mean cartilage thickness within a region of interest, denoted $ThCtAB$, as the weighted mean of cartilage thickness measurements at measurement points, weighting the contribution at each point by the sum of areas of its surrounding triangular facets. Points on the boundary of the region were included, but only facets within the region contributed to their weighting. For each functional sub-compartment, the mean cartilage thickness was computed over both subchondral ($ThCtAB$) and trimmed ($ThC^{Trm}AB$) coverage regions.

3 Results

MDL statistical shape models, consisting of 4098 points each, were built for the femur tibia and patella bones. The patella model's 4098 points provided mean connected point separation of 1.09mm. To ensure sufficient sampling of cartilage thickness, linear interpolation of the femur and tibia model points was performed which resulted in 16386 points with mean separations of 1.32mm and 1.07mm respectively.

Figure 1 shows the aggregate coverage map, shown on the mean bone shapes, and the subchondral and trimmed regions. The subchondral area allows compartmental volumes to be normalised for inter-subject variation in joint sizes, and provides anatomically consistent regions over which to compute mean thickness for each individual. The trimmed regions, which are completely encompassed by the subchondral region, allow mean thickness to be measured over the central, load-bearing regions of the joint. This excludes the edges of the cartilage which are more prone to segmentation errors and less involved in the disease process.

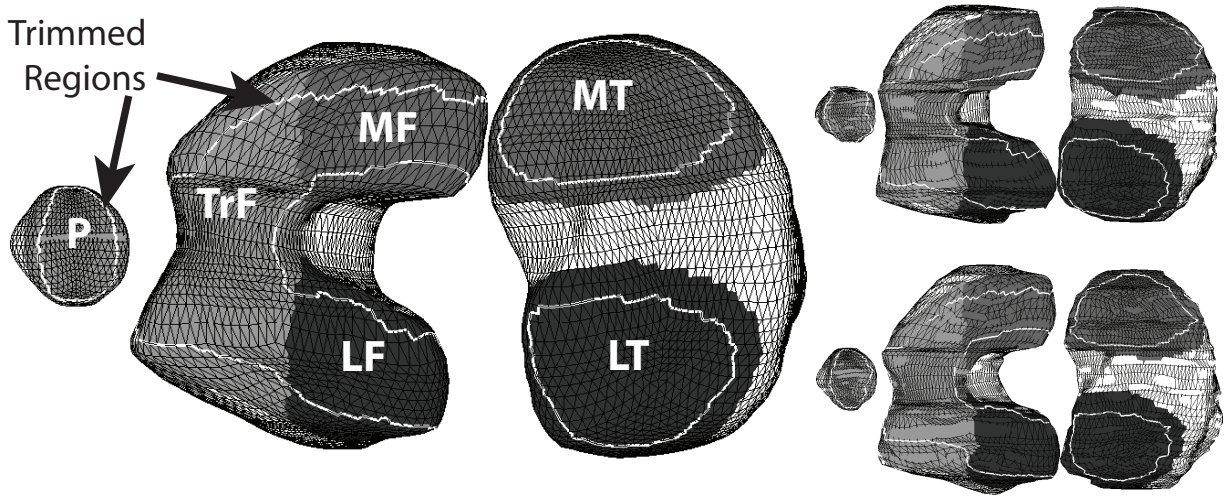


Figure 2. Regions of interest defined on the mean bone shapes (top) and propagated to two of the individuals' bone shapes. P: Patella, TrF: Trochlearfemur, MF: Medial Femur LF: Lateral Femur, MT: Medial Tibia LT: Lateral Tibia.

Figure 2 shows the functional regions drawn on the mean bone shapes and their propagation onto examples of the individual bones. Also shown are the trimming boundaries that allow measurements restricted to the central, load-bearing regions of the joint. The anatomical integrity of the regions are largely maintained when propagated onto the individual bone shapes.

The intra-segmentor and inter-subject variability of mean thickness measures is shown in Table 1. They demonstrate that repeatability is improved when measurement is restricted to the trimmed regions, implying that measurements taken at the edges of the cartilage are less reliable [13].

Morphological	Intra-Segmentor				Inter-Subject			
Measure CoV(%)	F	P	LT	MT	F	P	LT	MT
VC/tAB	2.91	3.18	3.23	4.80	13.8	15.4	23.6	15.9
$ThCtAB$	2.86	2.91	3.26	4.89	13.4	16.0	24.0	15.9
$ThC^{Trm}AB$	2.79	2.58	2.91	4.04	12.6	17.1	17.9	14.6

Table 1. Intra-Segmenter and Inter-Subject coefficient of variation.

Table 2 presents the normal range of mean cartilage thickness for all functional regions derived from the 19 healthy female volunteers. The results indicate that cartilage is thicker in the central, load bearing regions. Cartilage is thicker in the patella, femoral patella groove and the lateral tibia which is consistent with others' published results [2, 6].

(mm)	F	TrF	LF	MF	P	LT	MT
VC/tAB	2.19 (0.29)	-	-	-	2.77 (0.39)	2.21 (0.32)	1.64 (0.17)
$ThCtAB$	1.19 (0.14)	1.35 (0.22)	1.06 (0.18)	1.14 (0.22)	1.89 (0.30)	1.60 (0.35)	1.36 (0.21)
$ThC^{Trm}AB$	1.90 (0.22)	2.00 (0.28)	1.80 (0.23)	1.87 (0.28)	2.82 (0.48)	2.54 (0.44)	1.91 (0.27)

Table 2. Normal range, mean (standard deviation), of mean thickness measures.

4 Discussion

We have introduced a method that defines dense, anatomical correspondences invariant to individual joint shape or size - an improvement on current coordinate frame alignment aggregation methods. By choosing the underlying bone, rather than the cartilage itself, as the frame of reference, the approach can cope with morphological and topological changes in the cartilage.

The method offers advantages in precision and operational efficiency by allowing regions of interest to be defined on the mean bone shapes, and propagated to individuals in an anatomically consistent manner. Regional analysis allows changes in cartilage thickness to be detected independently in functional sub-regions, and also allows analysis to be restricted to the central, load-bearing regions of the joint, where change is more likely to be detected. The intra-segmentor coefficients of variation for the mean thickness measurements for the femur, patella, lateral tibia and medial

tibia compartments were reduced from 2.86% to 2.79%, 2.91% to 2.58%, 3.26% to 2.91% and 4.89% to 4.04% when measurement was restricted to central (trimmed) regions. This suggests that segmentation of the central thickness of the cartilage is more repeatable than the periphery of the cartilage sheet. Similarly, inter-subject variation for the femur, lateral tibia and medial tibia compartments decreased from 13.4% to 12.6%, 24.0% to 15.9% and 17.9% to 14.6% demonstrating less inter-subject variability in the central regions of the joint making the measures more sensitive, and increasing the likelihood of detecting change in cohort studies.

Acknowledgments

We acknowledge the study investigators S. J. Gandy, R. J. Moots, P. Creamer and A. J. Silman and Imophics Ltd. Manchester for the use of their medical image analysis software.

References

1. P. G. Bullough. "Osteoarthritis and related disorders: pathology." In J. H. Klippel, P. A. Dieppe & F. C. Arnett (editors), *Rheumatology*, volume 2, pp. 8.8.1–8.8.8. Mosby Yearbook Inc., St Louis, MO, second edition, 1988.
2. F. Eckstein & C. Glaser. "Measuring cartilage morphology with quantitative magnetic resonance imaging." *Seminars in Musculoskeletal Radiology* **8(4)**, pp. 329–353, 2004.
3. P. A. Hardy, P. Nammalwar & S. Kuo. "Measuring the thickness of articular cartilage from MR images." *J. Magnetic Resonance Imaging* **13**, pp. 120–126, 2001.
4. F. Eckstein, M. Hudelmaier, W. Wirth et al. "Double echo steady state (dess) magnetic resonance imaging of knee articular cartilage at 3 tesla - a pilot study for the osteoarthritis initiative." *Ann Rheum Dis* p. ard.2005.039370, 2005.
5. G. Li, S. E. Park, L. E. DeFrate et al. "The cartilage thickness distribution in the tibiofemoral joint and its correlation with cartilage-to-cartilage contact." *Clinical Biomechanics* **20(7)**, pp. 736–744, 2005.
6. Z. A. Cohen, V. C. Mow, J. H. Henry et al. "Templates of the cartilage layers of the patellofemoral joint and their use in the assessment of osteoarthritic cartilage damage." *Osteoarthritis and Cartilage* **11**, pp. 569–579, 2003.
7. S. R. Morgan, A. F. P. Nash, J. C. Waterton et al. "Magnetic resonance imaging measurement of knee cartilage volume in a multi-centre study." *Rheumatology* **42**, pp. 1–4, 2003.
8. T. G. Williams, C. J. Taylor, J. C. Waterton et al. "Population analysis of knee cartilage thickness maps using model based correspondences." In *ISBI 2004 2004 IEEE Int. Symp. on Biomedical Imaging*, pp. 193–196. IEEE, Arlington, VA, USA., 2004.
9. R. H. Davies, C. J. Twining, T. F. Cootes et al. "3D statistical shape models using direct optimisation of description length." In *ECCV 2002: 7th European Conference on Computer Vision, LNCS 2352/2002*, volume III, pp. 3–21. Springer, Copenhagen, Denmark, 2002.
10. F. Eckstein, G. Ateshian, R. Burgkart et al. "Proposal for a nomenclature for magnetic resonance imaging based measures of articular cartilage in osteoarthritis." *Osteoarthritis and Cartilage* **(to appear)**, 2006.
11. S. W. Hughes, T. J. DArcy, D. J. Maxwell et al. "Application of a new discreet *(sic)* form of gauss' theorem for measuring volume." *Phys. Med. Biol.* **41**, pp. 1809–1821, 1996.
12. R. Burgkart, C. Glaser, S. Hinterwimmer et al. "Feasibility of T and Z scores from magnetic resonance imaging data for quantification of cartilage loss in osteoarthritis." *Arthritis and Rheumatism* **48(10)**, pp. 2829–35, 2003.
13. S. Koo, G. E. Gold & T. P. Andriacchi. "Considerations in measuring cartilage thickness using MRI: factors influencing reproducibility and accuracy." *Osteoarthritis and Cartilage* **13(9)**, pp. 782–789, 2005.

Automatic segmentation of lumbar vertebrae on digitised radiographs using linked active appearance models

M.G. Roberts[a], T.F. Cootes[a], J.E. Adams[a]

[a]Department of Imaging Science and Biomedical Engineering
University of Manchester
Manchester, M13 9PL, UK
*

Abstract. Manual point placement for vertebral morphometry is time-consuming and imprecise, and morphometric methods for vertebral fracture diagnosis are unreliable. Automatic computer determination of the detailed vertebral shape could enable more powerful quantitative classifiers of osteoporotic vertebral fracture. The shape and appearance of vertebrae on 250 digitised lumbar radiographs were statistically modelled, using a sequence of active appearance models (AAMs) of overlapping triplets of vertebrae. To automatically locate the vertebrae, the sequence of models was matched to previously unseen scans. Accuracy results (0.64mm mean point-to-line error) were found to be similar to previously published results for dual-energy X-ray absorptiometry (DXA), but a low fracture prevalence meant that the shape models were undertrained for the few moderate and severe fractures. However mild fractures were fitted with good accuracy (mean 0.84mm). The results confirm the feasibility of substantially automating vertebral morphometry measurements on radiographs, despite the projective effects of the divergent X-ray beam. Use of the shape and appearance parameters of the models could in future provide a quantified form of some of the more subtle aspects of visual or semi-quantitative expert reading of vertebral fractures.

1 Introduction

Osteoporosis is a progressive skeletal disease characterised by a reduction in bone mass, resulting in an increased risk of fractures. Vertebral fractures are the most common, and occur in younger patients. The presence of vertebral fractures significantly increases the risk of further vertebral and non-vertebral fractures [1, 2]. The accurate identification of prevalent vertebral fractures is therefore clinically important, and the detection of incident vertebral fractures is important in evaluating new osteoporosis therapies. However there is no precise definition of exactly what constitutes a vertebral fracture, though a variety of methods of describing them have been developed [3]. These include semi-quantitative methods [4] involving some subjective judgement by an expert radiologist, and fully quantitative morphometric methods [5, 6]. The latter require the manual annotation of six (or more) points on each vertebra. This annotation is time consuming, and subtle shape information is lost in the reduction of shape to 6 points. Our ultimate aim is to define more reliable quantitative fracture classification methods based on a complete definition of the vertebra's shape and the surrounding texture. The first step must therefore be to achieve a reliable automatic segmentation. Some success in automatically locating vertebrae has been reported by several authors [7–10]. Most previous work on model-based vision applied to vertebrae has used DXA images, though [9, 10] used lumbar radiographs. DXA has many advantages, such as low radiation dose, and a lack of projective effects, but despite the growing use of DXA for vertebral fracture assessment, radiographs (or computed radiography) remain the definitive means of diagnosis.

We have already developed Active Appearance Model (AAM) [11, 12] based algorithms for segmenting vertebrae, and successfully applied these to DXA images [8, 13]. The purpose of this study was to assess the accuracy of this approach to segmenting vertebrae on radiographs. Although radiographs typically have better resolution and signal to noise ratio, the shape and appearance of the vertebrae is more complex due to projectional parallax effects. The divergent beam used in conventional radiography causes a variable scaling across the image, and can cause severe apparent tilting of the vertebral bodies. Also as the more extreme vertebral bodies tend to be obliquely irradiated, their superior and inferior endplates typically appear as elliptical rims, rather than the more linear edge typical of DXA. Figure 1 shows a typical lumbar radiograph, with some contrast enhancement to ensure all vertebrae are simultaneously visible.

2 Materials and Methods

2.1 Data

The images used were obtained from radiographs collected in a previous epidemiological study. We have thoracic and lumbar radiographs, but have initially just used lumbar radiographs as these are the more straightforward case

*martin.roberts@manchester.ac.uk

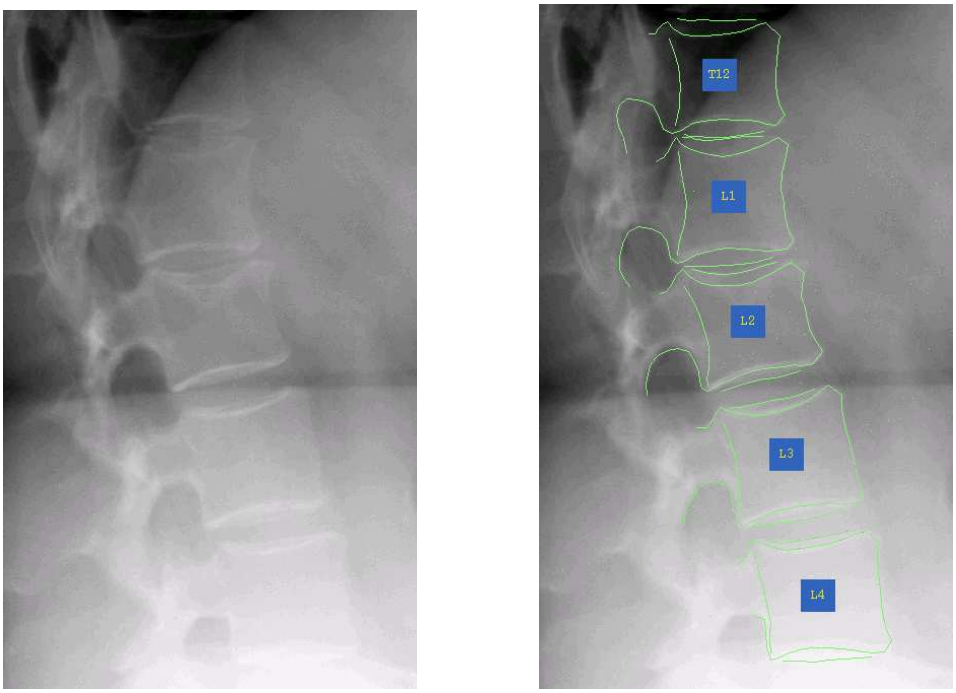


Figure 1. Lumbar radiograph. a) shows the raw image (contrast enhanced); b) shows the automatically located vertebral contours superimposed.

due to less clutter (e.g. from lungs and ribs), and a lower fracture prevalence. The dataset consisted of 250 lumbar radiographs, digitised using a Vidar [1] Diagnostic Pro Advantage digitiser at 300dpi and 12 bit intensity resolution. This Vidar digitiser allows a variety of analogue to digital conversion mappings. As there is typically a large range of brightness/contrast at different vertebral levels in the radiographs it was important to select a transform that preserved information across a large dynamic range. The default logarithmic transform did not work well on these images, as it typically "washed out" the often brighter vertebrae in the lower lumbar, whereas using a more nearly linear transform had the opposite effect of losing information in the typically darker upper portion (T12/L1). After some initial experimentation it appeared that the "power 3" [2] option gave the best compromise performance.

The digitised images were manually annotated using an in-house tool, by an experienced radiograper, supervised by the first author. Each vertebral contour uses 60 points around the vertebral body with 8 further points around the pedicles. The endplate rims were modelled using a quasi-elliptical shape, rather than the single edge previously used for DXA images. No images were included where the projectively induced tilting was so severe that lumbar vertebrae appeared to interpenetrate each other (with the occasional exception of the extreme T12/L1 or L4/L5 pairs). Such images essentially represent a setup error, and are extremely difficult to read, even by an expert radiologist, and lead to unreliable diagnosis. Figure 2 shows a zoomed in view of L3 with its shape model points displayed.

2.2 AAM approach

The method of [13] was used to fit a sequence of three AAMs composed of overlapping vertebral triplets covering the spine from L4 up to T12. Note that L5 is not normally used in vertebral fracture assessment as it is very rare for L5 to suffer osteoporotic fracture, and it may be obscured by the iliac crest. The three triplet models used were T12/L1/L2, L1/L2/L3 and L2/L3/L4. Results on DXA [8, 13] indicate that triplets of vertebrae are the optimal structure to model. If the modelled structure is too large then it is more prone to undertraining problems and latent non-linearities, and does not cope with local displacement pathologies such as scoliosis; whereas a smaller structure is too unconstrained. The algorithm of [8] combines the results of multiple sub-models with overlapping regions by fitting the models in a sequence, and using the constrained AAM [14] with constraints associated with the overlapping points. The extension to this algorithm in [13] used a "best-fit first" heuristic to obtain a close-to-optimal ordering of the model fitting sequence. In effect each (non-extreme) vertebrae is fitted using the triplet sub-model in which it is central, and the

[1] Vidar Systems Corp, Herndon VA, USA
[2] manufacturer's designation, in fact it appears to be a cube root

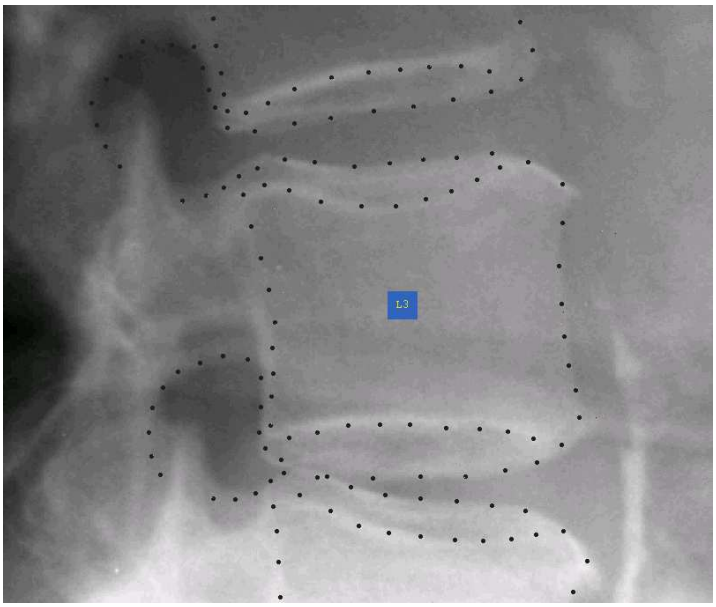


Figure 2. Zoomed in view of L3 showing its shape model points

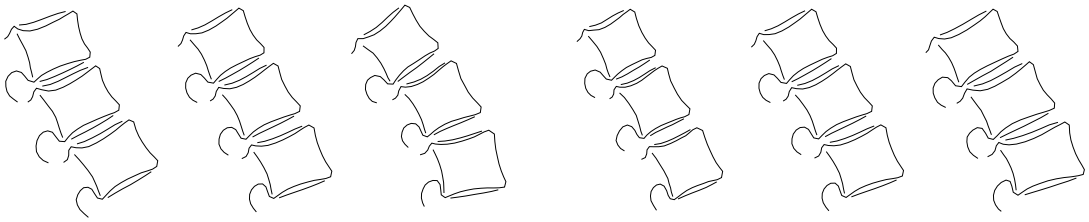


Figure 3. L2 triplet 3SD variation in first (left) and second (right)shape modes

neighbours provide helpful constraints and linkage to the other sub-models. See [8] for details. Each triplet sub-model has its own affine pose parameters. Figure 3 shows the variation in the first two shape modes of the L2-centred triplet.

T12 was included to form the uppermost L1-centred triplet, although there were some lumbar radiographs in which T12 was not fully visible. Nevertheless in general T12 should be visible on a lumbar radiograph, and results from DXA lead us to believe that it is helpful in fitting L1 to also include the neighbouring T12 in the model. In fact sometimes T12 is better visualised on the lumbar radiograph than on the thoracic. There was often a high variation in brightness and contrast across the different vertebral levels. For example T12 or even L1 were often very dark, and could often not be seen without some local contrast optimisation, whereas L4 typically had an over-bright "washed-out" appearance. Figure 1 is typical in this respect. Another advantage of decomposing the overall shape into sub-structures is that the texture normalisation can be better tuned to the local brightness and contrast, where there is substantial variation in these across the image.

As there is little useful information inside the vertebral body we used profile samplers for the AAM texture model, rather than the triangulated region samplers classically used with an AAM. The profile samplers extracted the gradient perpendicular to the local shape, and this was non-linearly renormalised using a sigmoidal function tuned to the mean absolute gradient [15] over the entire profile set. We used a 4-level multi-resolution pyramid search, to extend the convergence zone, with 8 samples either side of the shape. The finest level step size was 0.375mm, and the images were pre-smoothed up to a resolution of 0.1694 mm per pixel (i.e. one level of Gaussian pyramid up-smoothing). Thus the profile step size represents about 2 pixels (at each level of the pyramid). The extracted gradient is Gaussian smoothed across the local tangent, with a smoothing window equal to the step length. We also experimented with a profile sampler which concatenated this profile with a similar profile sampler extracting a measure of image corner strength

	Normal				Fractured			
Profile Sampler	Mean Acc	Median Acc	75%ile Acc	%ge errors over 2mm	Mean Acc	Median Acc	75%ile Acc	%ge errors over 2mm
Gradient Only	0.71	0.46	0.89	6.2%	1.11	0.62	1.34	14.1%
Gradient & Corner	0.64	0.43	0.82	4.6%	1.06	0.61	1.32	13.3%

Table 1. Search Accuracy Percentiles by Fracture Status for the two profile samplers used

("cornerness"), as in [15]. As the corners of the vertebrae are of physical interest in standard morphometry, it was thought that including a cornerness measure in the AAM might improve the accuracy at points of important diagnostic interest. Furthermore the projective parallax and oblique beam orientation tend to introduce curved features in the region of the profile. The cornerness measure has the further advantage that it implicity includes feature information from a somewhat larger region, as the measure is based on the structure tensor $(\nabla I \nabla I^T)$ [3], which is Gaussian smoothed over a square region with semi-width twice the profile step length. See [15] for details.

2.3 Experiments

Leave-25-out tests were performed over the 250 images. As AAMs perform local search an approximate initialisation somewhere in the vicinity of the vertebrae is needed. When the algorithm is run interactively in an associated prototype clinical tool, the clinician initialises the solution by clicking on the appoximate centres of each vertebra. A global shape model is then used to give an approximate starting solution with best least squares fit of its vertebral centres to the input points. The same method was used in this study, and on each experiment the user imprecision was simulated by using the known equivalent marked points and adding random offsets to them. These were zero-mean Gaussian errors with SD of 2mm in the y-direction (along the spine) and 3mm in the x-direction. Twenty replications (i.e. random initialisations) of each image were performed.

3 Results

The accuracy of the search was characterised by calculating the absolute point-to-line distance error for each point on the vertebral body. Table 1 compares results for the two profile samplers used with the data separated into points within normal or fractured vertebrae. Each row gives the mean, median and 75th percentiles, and the percentage of point errors in excess of 2mm. The threshold of 2mm would be around 2.5 SDs of manual precision, and can be viewed as a point failure indicator.

Table 1 shows that the results are worse for fractured than normal vertebrae. A more detailed examination by fracture grade gives mean accuracies of 0.84mm, 1.79mm and 3.35mm for fracture grades 1, 2 and 3 [4] respectively (for the gradient and corner sampler). However there was a low fracture prevalence in the lumbar region in the sample, and these figures are based on 17 grade 1 fractured vertebrae, 2 grade 2 and only a single grade 3 fracture.

The more sophisticated profile sampler including a corner measure appears to produce a small improvement in accuracy of around 0.07mm. This represents a 10% reduction in mean error. We confirmed that this difference is statistically significant at the 1% level by bootstrap resampling of the differences in errors between the two profiles. This enables the derivation of a symmetric (in probability) 99% bootstrapped confidence interval on the mean difference of [0.048,0.082]. As this interval does not span zero, the difference is significant at the 1% level.

4 Discussion

4.1 Overall Accuracy Performance

The mean segmentation accuracy of 0.64mm on normal vertebrae is comparable to manual precision in point placement, and to previous results on DXA images [13]. Over 95% of points in normal vertebrae are located to within 2mm of the manually annotated outline. However the dataset contained a very low prevalence of fractured vertebrae, and so the shape models are evidently undertrained, for fractures above grade 1. Therefore within the limitations of the small fractured sample it appears that the results deteriorate with increasing fracture grade. However given previous reasonable accuracy achieved on fractured vertebrae with DXA images [13], we believe that this problem could be

[3] with the Cartesian image gradient ∇I as a column vector

[4] i.e. mild, moderate and severe fractures, see [4]

solved by adding more fractured training examples. The mean accuracy is better than other comparable cited figures in the literature [9, 10]. For example de Bruijne *et al* [10] obtained a mean point-to-contour accuracy of 1.4mm on lumbar radiographs using shape particle filtering, which is more than double the size of error achieved by our AAM approach. On the other hand this was for a fully automatic search with no approximate manual initialisation such as we use. Howe *et al* [9] state that 68% of points on lumbar radiographs were located to within 25 pixels. We understand the dataset used had a resolution of 0.174mm per pixel, so this is equivalent to a 68th percentile of 4.35mm, clearly substantially worse than our 75th percentile of 0.85mm. However again Howe *et al* were using a completely automatic method, with the AAM being initialised to the best template match found by an initial Generalised Hough Transform.

4.2 Conclusion

In conclusion the results confirm the feasibility of substantially automating vertebral morphometry measurements on radiographs, although the shape models need better training on fractured vertebrae. Within the limitations of the dataset, the projective effects of spinal radiography do not appear to present any substantial problem to an AAM-based approach.

4.3 Future Work

We intend to extend the work to the thoracic spine, which tends to contain more osteoporotic fractures. Use of the shape and appearance parameters of the fitted models could in future provide a means of classifying vertebrae as normal, fractured , or otherwise deformed. Current simplistic quantitative morphometric methods are unreliable, especially for mild fractures, but the appearance parameters may provide a quantified form of some of the more subtle aspects of visual or semi-quantitative expert reading of vertebral fractures. We therefore view obtaining a reliable automatic segmentation as the first step in achieving a Computer Aided Diagnosis (CAD) system for vertebral fracture.

5 Acknowledgements

The authors would like to thank Stephen Capener who performed the manual annotation of the vertebrae.

References

1. D. Black, N. Arden, L. Palermo et al. "Prevalent vertebral deformities predict hip fractures and new vertebral deformities but not wrist fractures." *Study of Osteoporotic Fractures Research Group, J Bone Miner Res* **14**, pp. 821–828, 1999.
2. L. Melton, E. Atkinson, C. Cooper et al. "Vertebral fractures predict subsequent fractures." *Osteoporosis Int* **10**, pp. 214–221, 1999.
3. A. Guermazi, A. Mohr, M. Grigorian et al. "Identification of vertebral fractures in osteoporosis." *Seminars in Musculoskeletal Radiology* **6(3)**, pp. 241–252, 2002.
4. H. Genant, C. Wu, C. van Kuijk et al. "Vertebral fracture assessment using a semi-quantitative technique." *J Bone Miner Res* **8**, pp. 1137–1148, 1993.
5. R. Eastell, S. Cedel, H. Wahner et al. "Classification of vertebral fractures." *J Bone Miner Res* **6(3)**, pp. 207–215, 1991.
6. E. McCloskey, T. Spector, K. Eyres et al. "The assessment of vertebral deformity: a method for use in population studies and clinical trials." *Osteoporosis Int* **3**, pp. 138–147, 1993.
7. P. Smyth, C. Taylor & J. Adams. "Vertebral shape: automatic measurement with active shape models." *Radiology* **211**, pp. 571–578, 1999.
8. M. Roberts, T. Cootes & J. Adams. "Linking sequences of active appearance sub-models via constraints: an application in automated vertebral morphometry." In *14th British Machine Vision Conference*, pp. 349–358. 2003.
9. B. Howe, A. Gururajan, H. Sari-Sarraf et al. "Hierarchical segmentation of cervical and lumbar vertebrae using a customized generalized hough transform and extensions to active appearance models." In *Proc IEEE 6th SSIAI*, pp. 182–186. 2004.
10. M. de Bruijne & M. Nielsen. "Image segmentation by shape particle filtering." In *International Conference on Pattern Recognition*, pp. 722–725. IEEE Computer Society Press, 2004.
11. T. Cootes & C. Taylor. "Statistical models of appearance for medical image analysis and computer vision." *Proc SPIE Medical Imaging* **3**, pp. 138–147, 2001.
12. T. Cootes, G. Edwards & C. Taylor. "Active appearance models." In *5th European Conference on Computer Vision*, volume 2, pp. 484–498. Springer (Berlin), 1998.
13. M. Roberts, T. Cootes & J. Adams. "Vertebral shape: Automatic measurement with dynamically sequenced active appearance models." In *8th MICCAI Conference*, volume 2, pp. 733–740. 2005.
14. T. Cootes & C. Taylor. "Constrained active appearance models." In *8th International Conference on Computer Vision*, volume 1, pp. 748–754. IEEE Computer Society Press, July 2001.
15. I. Scott, T. Cootes & C. Taylor. "Improving active appearance model matching using local image structure." In *18th Conference on Information Processing in Medical Imaging*, pp. 258–269. 2003.

Quantifying Microvascular Brain Disease: Identification of Potential Imaging-Based Biomarkers

Alan Jackson

Imaging Science and Biomedical Engineering, University of Manchester

Microvascular brain disease is one of the commonest disorders to affect the aging population. In normal aging there is significant evidence of microvascular deterioration caused by aging and pressure related damage to the walls of the small arterial vessels supplying the brain. Atheromatous disease affects major vessels such as the carotid artery where it has been widely studied as a source of morbidity. The atheromatous process also affects the penetrating arterioles within the brain at a scale too small to identify on modern imaging. Despite this there is overwhelming evidence that microvascular disease is of very significant implication both in aging and in a range of vascular and age-related diseases including vascular dementia, Alzheimer's disease, late onset depression and others.

One of the major problems with microvascular brain disease is that it is impossible to definitively diagnose its presence until after death when histology can demonstrate the arteriolar walls and associated ischaemic brain injury. Identification of the presence and severity of microvascular brain disease during life is of major importance both for neuroscience research and for clinical practice. At the present time indicators include clinical indicators of vascular risk such as the presence of high blood pressure, smoking etc which are very non-specific indeed but do correlate with the presence of microvascular disease and the presence of white matter lesions in the brain on MR imaging. These lesions are non-specific evidence of vascular injury which are widely used as a direct surrogate indicator of the presence of microvascular brain disease. A number of semi-quantitative and quantitative methods for their assessment do exist. Unfortunately they are very much a secondary effect of microvascular brain disease and there is a need for direct quantitative biomarkers of the disorder. We will discuss the nature of microvascular brain disease and its clinical expression. We will then discuss the possibility of developing imaging-based biomarkers focussed on the anatomical and physiological features of the disorder. In particular we will address the visualisation of microvascular angiopathy, a complex and histological change in microvascular disease which includes dilatation of perivascular spaces which can be seen by high resolution magnetic resonance imaging. We will present studies of these features and a simple semi-quantitative scoring system in comparison with existing methodologies as a potential biomarker.

In addition we will discuss the potential use of physiological functional imaging to identify microvascular brain disease. This is based on a hypothesis that the presence of atheroma within small vessels gives rise to a decrease in small vessel compliance. We will present studies which will demonstrate secondary evidence of decreased vascular compliance using magnetic resonance based measurements of fluid pulsatility in the basal arteries and CSF spaces. These semi quantitative measures are shown to correlate with clinical indices of microvascular disease. There is also however a need for quantitative modelling of the physiological changes which occur in microvasclar disease and we will present a preliminary study of such a model and discuss its potential shortcomings and application to study the nature of microvascular disease.

Automated gamma knife radiosurgery treatment planning with image registration, data-mining, and Nelder-Mead simplex optimisation

Kuan J Lee[a*], David C Barber[b] and Lee Walton[bc].

[a]Academic Radiology, University of Sheffield, Sheffield, UK
[b]Medical Physics, Sheffield Teaching Hospitals Trust, Sheffield, UK
[c]The National Centre for Stereotactic Radiosurgery, Sheffield, UK

Abstract. We present a new, fully automatic method of gamma knife treatment planning. After using image registration to convert lesions into meshes, we data-mine the database of previously treated lesions to find a close shape match for the new target lesion. The treatment parameters in the previous plan are transferred to the new lesion, and computer optimised until a conformal dose distribution for the new target is found. The method was tested on a randomly selected set of 10 acoustic neuromas and 10 meningiomas. In 8/10 acoustic neuromas and 8/10 meningiomas, the automatic method found plans with conformation number equal or better than that of the manual plan. In 4/10 acoustic neuromas and 5/10 meningiomas, both overtreatment and undertreatment ratios were equal or better in automated plans. In the majority of the test set cases, the automated plans also gave shorter estimated treatment times.

1 Introduction

The Leksell Gamma Knife uses a focused array of 201 intersecting beams of ^{60}Co gamma radiation to treat lesions within the brain. The combined intensity of radiation at the focus (or isocentre) is extremely high while the intensity only a short distance from the isocentre is very low. Four collimator sizes are available (4, 8, 14 and 18 mm), but almost always, the dose from a single focus of radiation will not match the shape of the lesion to be treated. By overlapping multiple foci of radiation of varying sizes, the dose distribution can be fashioned to produce a conformal treatment plan. The main parameters in a plan are: the number of isocentres, their collimator sizes, relative weights, and position coordinates in 3 dimensions. The usual aim is a prescription isodose contour that is maximally conformal with the lesion outline. Because predicting the resulting dose distribution from the combination of several isocentres is non-intuitive, automation is extremely desirable.

To make the problem tractable, it is broken into two parts. A pre-optimisation stage provides an initial plan as a "starting point". This is followed by computer optimisation of some or all of the parameters. For example, Yan et al [1] and Shu et al [2] pre-optimise with a rough manual plan which is optimised with Powell's method, and quasi-Newton plus simulated annealing respectively. To reduce the dependence on human input, Leichtman et al specified only the maximum number of isocentres before optimising with simulated annealing [3]. The groups at Case Western University [4, 5] and University of Wisconson [6, 7] minimise human interaction even further by deriving their starting point using a geometry-based heuristic. They model the dose distributions from isocentre as spheres which are to be placed along a "skeleton" of the target volume, calculated from the medial axis transform. Through a series of packing steps, their algorithms determine the starting number of isocentres, locations, sizes and weights, which are then optimised using various methods, including linear programming, non-linear programming and evolutionary simulated annealing.

Our approach is to pre-optimise by data-mining the database of past plans with shape-matching. First, a close shape match between a previous lesion in a database of past treatment plans and the new lesion is identified. The treatment parameters in the previous plan are then used as a starting point for subsequent computer optimisation. To test the method, automatic plans were generated for a set of acoustic neuromas and meningiomas, and compared with the actual treatment plans used.

2 Methods

2.1 Database

1316 plans calculated for our Leksell Gamma Knife model C were extracted for inclusion in the database. All plans were reflected about the sagittal plane to exploit left-right symmetry of the system, giving a total number of 2632 plans.

*Email:k.j.lee@sheffield.ac.uk

2.2 Data-mining

2.2.1 Mesh representation of lesion

On the Leksell GammaPlan system, a lesion volume is contoured manually and becomes represented as a collection of 2d polygons with non-uniformly spaced, manually defined vertices in axial planes approximately 1 mm apart. The non-uniform spacing of vertices can bias lesion matching. Therefore we represent the lesion volume as a mesh of uniformly spaced nodes, by mapping a spherical mesh of uniformly spaced nodes so that its surface lies close to the manual points. We accomplish this using an image registration method based on gradients [8,9], but because the method works best between volumes of similar size, we split the mapping into three parts: (i) a mapping from a spherical mesh to an ellipsoidal mesh that "best fits" the manual points, which is used to map the manual points closer to the spherical mesh, (ii) a mapping from the transformed manual points to the spherical mesh, and (iii) applying to the spherical mesh the inverse mappings of part (ii) then (i) to transform the mesh back to the space of the original points. Details can be found in the Appendix.

2.2.2 Lesion matching

The target lesion and the database lesion being tested against are first converted to (binary) images and the exclusive-OR formed between these images. The resulting image represents those voxels which belong to at least one but not both images. The total number of such voxels is a volume. This is then divided by the surface area of the database lesion to give a length. The result, which we call the "Barber distance" b, is very fast to compute, and can be interpreted as a measure of the average distance between the surfaces of the two lesions.

Database lesion volumes with the shortest b to the target were extracted from the database together with the isocentre sizes, weights and positions previously used to treat the database lesion. These isocentre treatment coordinates were translated to the target lesion being matched (by matching centre of gravities of the two volumes). These new positions, sizes and weights constitute the starting parameters for the subsequent computer optimisation.

2.2.3 Nelder-Mead optimisation

The 3d coordinates and weights of the isocentres derived from the matching process were optimised using the Nelder-Mead simplex method [10]. To derive the points of the initial simplex, variables that were coordinates were perturbed by 5 mm, and weight variables were increased by 1.5 times.

The cost function CF was defined as:

$$CF = -\left(\frac{TV_{PIV}}{PIV}\right)\left(\frac{TV_{PIV}}{TV}\right)^{n} \tag{1}$$

where n is an exponent, TV is the target volume, PIV is the volume covered by the prescription isodose (taken to be 50%), and TV_{PIV} is the volume of the target covered by the prescription isodose. The first term on the right is an overtreatment ratio (OR), and the second term an undertreatment ratio (UR) [11]. The exponent alters their relative contributions to the cost function, so that $n > 1$ biases the optimisation towards a more generous plan. If $n = 1$, the cost function becomes equal in magnitude to the conformation number (CN) [11]. We used value of $n = 1.5$ in the cost function. This was determined experimentally from a small set of cases (none of which is in the test set used here), as being roughly the value which gave automatic plans that were most similar to the manual plans.

All dose points were calculated in full on grid points at 2 mm separation using the KULA (v3.1) treatment planning method (Elekta Instrument AB, Stockholm, Sweden). Skull shapes were approximated as ellipsoids, the dimensions and position of which were derived from the skull geometry ("bubble") measurements used in GammaPlan.

2.2.4 Test set

Ten acoustic neuromas and 10 meningiomas were randomly selected from new patients who attended for treatment after the formation of the lesion database. The only criterion was that the cases should not have used any plugging, as we have not yet included this factor. Table 1 shows the CN, OR and UR values from manual plans for acoustic neuromas, arranged in order of increasing number of shots N, which is an indication of the complexity of the planning task. The table also shows each plan's estimated time T for treatment delivery as calculated by the Leksell GammaPlan, with a 15 minute penalty added for every collimator change. Table 2 shows the corresponding values for the meningiomas.

For each target volume, the 12 best candidate matches were found and optimised as described. The CN, OR and UR values were calculated for all results and compared with corresponding values from the manual plan.

Patient	Target volume (mm^3)	N		T (mins)		CN		OR		UR	
		Man	Auto	Man	Auto	Man	Auto	Man	Auto	Man	Auto
1	1885	4	14	51	55	0.87	0.83	0.91	0.88	0.95	0.93
2	*716*	*4*	*6*	*44*	*38*	*0.76*	*0.81*	*0.86*	*0.89*	*0.89*	*0.91*
3	*180*	*4*	*5*	*30*	*19*	*0.70*	*0.74*	*0.79*	*0.81*	*0.89*	*0.92*
4	924	6	5	48	16	0.84	0.85	0.91	0.90	0.93	0.95
5	*3581*	*7*	*11*	*76*	*63*	*0.78*	*0.80*	*0.85*	*0.85*	*0.92*	*0.94*
6	1627	7	7	59	47	0.75	0.75	0.86	0.83	0.88	0.91
7	882	8	8	63	49	0.78	0.83	0.82	0.90	0.96	0.92
8	10960	12	11	65	48	0.87	0.83	0.96	0.91	0.91	0.91
9	1611	12	8	63	43	0.81	0.81	0.90	0.88	0.91	0.92
10	*4099*	*15*	*18*	*98*	*59*	*0.67*	*0.72*	*0.80*	*0.82*	*0.83*	*0.87*

Table 1. Manual vs. automatic plans for 10 acoustic neuromas

Patient	Target volume (mm^3)	N		T (mins)		CN		OR		UR	
		Man	Auto	Man	Auto	Man	Auto	Man	Auto	Man	Auto
1	421	1	9	21	70	0.70	0.79	0.73	0.88	0.95	0.90
2	2105	4	10	61	78	0.65	0.79	0.65	0.86	0.99	0.92
3	12196	8	12	36	54	0.78	0.81	0.92	0.90	0.85	0.90
4	*16891*	*8*	*10*	*57*	*77*	*0.67*	*0.74*	*0.84*	*0.86*	*0.80*	*0.85*
5	3325	9	10	75	70	0.69	0.65	0.83	0.74	0.83	0.87
6	*5452*	*11*	*15*	*71*	*91*	*0.69*	*0.74*	*0.81*	*0.83*	*0.85*	*0.89*
7	*3324*	*12*	*12*	*89*	*62*	*0.72*	*0.79*	*0.87*	*0.87*	*0.83*	*0.91*
8	*3982*	*15*	*14*	*92*	*77*	*0.56*	*0.63*	*0.66*	*0.71*	*0.84*	*0.89*
9	37300	16	9	88	40	0.64	0.60	0.73	0.73	0.88	0.82
10	*12962*	*22*	*8*	*115*	*55*	*0.41*	*0.44*	*0.58*	*0.62*	*0.69*	*0.71*

Table 2. Manual vs. automatic plans for 10 meningiomas

3 Results

3.1 Lesion matching

Matching a target against 2632 database volumes with the Barber distance b took under 30 s seconds on a 2 GHz processor. However, the best automated plan was not always obtained with the lowest b. Overall, the average rank (as defined by b) of the match that gave the best automated plan was 4.7, with range 1-12.

3.2 Automated plans

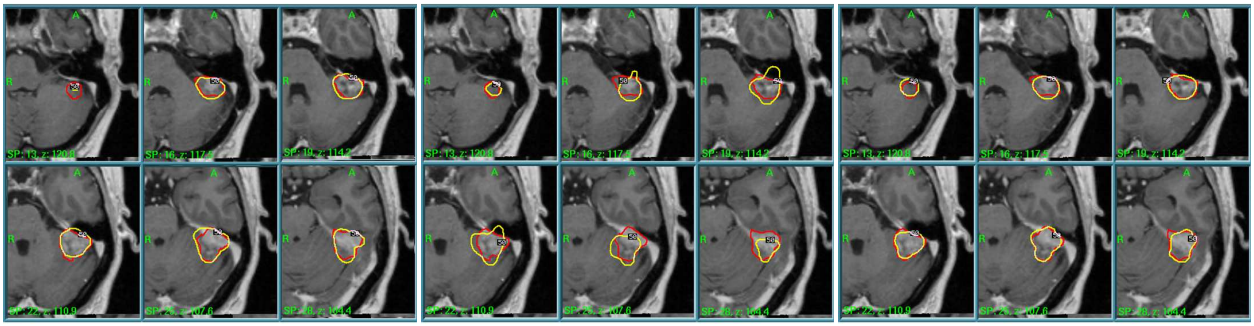
Table 1 shows that of 8/10 acoustic neuromas, automated plans gave the same or better CN. Of these, 4 automated plans had better values of both OR and UR - these are shown in italics. In 9/10 cases, the automated plans also resulted in a shorter estimated treatment time.

Table 2 shows that in 8/10 meningiomas, automated plans gave the same or better CN. Of these, 5 automated plans had better values of both OR and UR (italics). Automated plans gave improved treatment times in 5/10 cases.

Representative slices from an example manual plan are shown in Fig 1(a). The same slices from the initial plan found by data-mining are shown in Fig 1(b), and those from the final automated plan after optimisation in Fig 1(c). In this case, the automated plan had superior measures (see Table 2).

4 Discussion

In the majority of cases considered, automated plans had better values of CN than manual plans. In some cases, it has achieved the difficult task of returning plans with better values of both OR and UR, often with also a reduced



(a) manual (b) match before optimisation (c) match after optimisation

Figure 1. Prescriptive isodose (yellow) and lesion (red) contours for meningioma patient number 6.

estimated treatment time. Nevertheless, there are many weaknesses in our current implementation, which need to be addressed before it can be used routinely.

On average, the best optimised plan was given by the match ranked fifth with the Barber distance. Future work will aim at finding a method of matching from which the best match will also give the best automated plan. A more complex matching metric may be needed. However, shape alone may not be enough: other factors e.g. number and position of isocentres, relative proportions of collimator sizes etc, may determine how flexibly one plan can be modified for use with another target. In this case, artificial intelligence methods might help. For instance, a set of target volumes and their best candidates could be used as a training set for computer learning e.g. with neural networks.

For optimisation, we chose the Nelder-Mead simplex method. Although optimisation took under a minute for small volumes ($< 1000\text{mm}^3$, 2-3 isocentres); for large volumes and number of isocentres ($> 30,000\text{mm}^3$, > 20 isocentres), one iteration could take 20 s to compute on a 3 GHz processor. With optimisation taking up to 5000 function evaluations, this gave a total optimisation time of tens of hours, which is unacceptable for routine use. To speed up optimisation, future work will examine the use of dose models [12], especially for large volumes where accurate dose distribution is relatively less important. Parallel computing and other methods of optimisation, such as those used by the above-mentioned groups, will also be explored.

In spite of all the difficulties, many aspects of our method make it worth investigating further. Data-mining provides an alternative method of generating initial plans for manual or computer optimisation. Moreover, as these plans already reflect the historic practice of the centre, this may have the advantage of producing plans more in keeping with what is considered acceptable by the planning team. Data-mining could also be a valuable training aid. Of course, a disadvantage is that better plans not present in the database (for example if the lesion volume is substantially different to anything treated before) will not be found. Until the database is sufficiently large and varied, there will be a need for introducing "new stock". One way is simply to add plans from other centres to the database.

References

1. Y. L. Yan, H. Z. Shu, X. D. Bao et al. "Clinical treatment planning optimization by Powell's method for gamma unit treatment system." *International Journal of Radiation Oncology Biology Physics* **39(1)**, pp. 247–254, 1997.
2. H. Z. Shu, Y. L. Yan, X. D. Bao et al. "Treatment planning optimization by quasi-newton and simulated annealing methods for gamma unit treatment system." *Physics in Medicine and Biology* **43(10)**, pp. 2795–2805, 1998.
3. G. S. Leichtman, A. L. Aita & H. W. Goldman. "Automated gamma knife dose planning using polygon clipping and adaptive simulated annealing." *Medical Physics* **27(1)**, pp. 154–162, 2000.
4. Q. J. Wu, V. Chankong, S. Jitprapaikulsarn et al. "Real-time inverse planning for gamma knife radiosurgery." *Medical Physics* **30(11)**, pp. 2988–2995, 2003.
5. P. P. Zhang, J. Wu, D. Dean et al. "Plug pattern optimization for gamma knife radiosurgery treatment planning." *International Journal of Radiation Oncology Biology Physics* **55(2)**, pp. 420–427, 2003.
6. M. C. Ferris, J. H. Lim & D. M. Shepard. "An optimization approach for radiosurgery treatment planning." *Siam Journal on Optimization* **13(3)**, pp. 921–937, 2003.
7. D. M. Shepard, L. S. Chin, S. J. DiBiase et al. "Clinical implementation of an automated planning system for gamma knife radiosurgery." *International Journal of Radiation Oncology Biology Physics* **56(5)**, pp. 1488–1494, 2003.
8. D. C. Barber. "Registration of low resolution medical images." *Physics in Medicine and Biology* **37(7)**, pp. 1485–1498, 1992.
9. D. C. Barber & D. R. Hose. "Automatic segmentation of medical images using image registration: diagnostic and simulation applications." *Journal of Medical Engineering & Technology* **29(2)**, pp. 53–63, 2005.

10. W. Press, S. Teukolsky, W. Vetterling et al. *Numerical Recipes in C*. Cambridge University Press, second edition, 1992.
11. I. Paddick. "A simple scoring ratio to index the conformity of radiosurgical treatment plans. technical note." *Journal of Neurosurgery* **93 Suppl 3**, pp. 219–22, 2000.
12. P. S. Cho, H. G. Kuterdem & n. Marks, R. J. "A spherical dose model for radiosurgery plan optimization." *Physics in Medicine and Biology* **43(10)**, pp. 3145–8, 1998.

Appendix

An ellipsoid has the general form:

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = r^2 \tag{2}$$

where $\mathbf{x}$ is a vector of co-ordinates lying on an ellipsoidal surface specified by the constant r and the matrix $\mathbf{A}$. To find the "best fit" to the given boundary points, a starting value for $\mathbf{A}$ is the inverse of the covariance matrix of the data points. The value of r^2 is determined by inserting the actual manual data points in turn into this equation, computing and averaging the values obtained. The matrix $\mathbf{B} = \mathbf{A}/r^2$ is formed. Then

$$\mathbf{x}^T\mathbf{B}\mathbf{x} = 1 \tag{3}$$

A spherical mesh of unit radius is generated using ANSYS (ANSYS, Inc. Canonsburg, PA, USA). For this mesh, for nodes $\mathbf{y}$ on the surface of this mesh,

$$\mathbf{y}^T\mathbf{y} = 1 \tag{4}$$

Equating these equations:

$$\mathbf{x}^T\mathbf{B}\mathbf{x} = \mathbf{y}^T\mathbf{y} \tag{5}$$

The singular value decomposition (SVD) of $\mathbf{B}$ is given by:

$$\mathbf{B} = \mathbf{U}\Lambda\mathbf{U}^T = \mathbf{U}\Lambda^{½}\Lambda^{½}\mathbf{U}^T \tag{6}$$

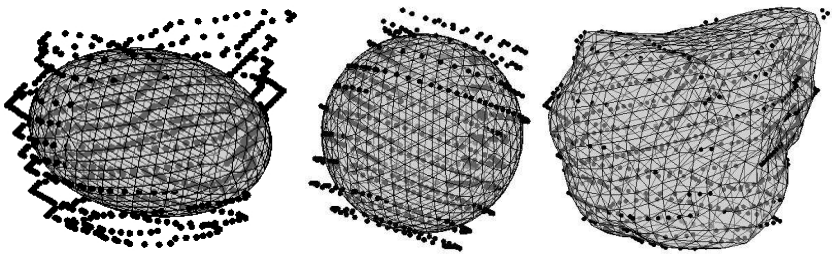
using which we derive from Eq. 5:

$$\mathbf{x}^T\mathbf{U}\Lambda^{½} = \mathbf{y}^T \tag{7}$$

and

$$\mathbf{x}^T = \mathbf{y}^T\Lambda^{-½}\mathbf{U}^T \tag{8}$$

This allows the nodes of the standard spherical mesh to be transformed into the nodes of an ellipsoidal mesh, the surface of which best fits the manual points (Fig 2(a)).



(a) (b) (c)

Figure 2. (a) Manual points with best-fit ellipsoid (b) manual points after applying same mapping as that of ellipsoidal mesh to spherical mesh (c) the spherical mesh has been mapped to lie close to the original manual points.

Next, both mesh and manual data points are converted to images in an image volume. An affine mapping is generated from the nodes of the ellipsoidal mesh and the equivalent nodes of the reference spherical mesh. This mapping is then applied to the manual points to map them closer to the reference sphere (Fig 2(b))

The resulting image after all the manual points have been so mapped is an image representing (incompletely) the surface of the lesion. The lesion surface image is then non-linearly registered to the reference sphere surface [9]. The result of this mapping is that the points representing the original lesion surface are now situated on the surface of the reference sphere. The inverse of this mapping is applied to the reference sphere to map the sphere to the mapped manual points. The affine mapping is then applied to the non-linearly mapped spherical mesh which maps this mesh back to the lesion position and shape (Fig 2(c)). The resulting mesh is now a representation of the lesion volume. The original manual points will now lie close to the surface of the mapped mesh. The original surface points do not correspond to any particular surface nodes of this mesh. The mesh is a representation of the lesion volume derived from (possibly incomplete) data available about the position of the lesion surface.

Thickness dependent tortuosity estimation for retinal blood vessels

Hind Azegrouz [a], Emanuele Trucco [a], Baljean Dhillon [b], Thomas MacGillivray [c] and I.J. MacCormick [d] *

[a]School of Engineering and Physical Sciences
Department of Electrical Electronic and Computer engineering
Heriot Watt University, EH14 4AS Riccarton, UK
[b]Princess Alexandra Eye Pavilion, Chalmers Street, Edinburgh, UK
[c]The Wellcome Trust Clinical Research Facility, Western General Hospital,
Crewe Road, Edinburgh, UK
[d]The College of Medicine and Veterinary Medicine, University of Edinburgh,
The Chancellor's Building 49 Little France Crescent, Edinburgh, UK

Abstract. This paper describes a framework for the automated estimation of vessel tortuosity in retinal images. We introduce a new tortuosity metric that takes into account vessel thickness, yielding estimates closer to intuition and medical judgement than those from previous metrics. We also propose an algorithm identifying automatically a vasculature segment connecting two points specified manually. Starting from a binary image of the vasculature, the algorithm computes a skeletal (medial axis) representation on which all terminal and branching points are located. This is then converted to a graph representation including connectivity as well as thickness information for all vessels. Target segments for tortuosity estimation are identified automatically from end points selected manually using a shortest-path algorithm. Results are presented and compared with those provided by clinical classification on 50 vessels from DRIVE images. An overall performance ratio of 92.4% is achieved.

1 INTRODUCTION

The tortuosity of blood vessels in the human retina is a risk indicator linked to many diseases, such as atherosclerosis and hypertension. A number of tortuosity measures have been proposed [1–4]. A common one is the *distance metric* [2] defined as the ratio of the vessel length to the chord length between the end points. Other measures proposed are based on curvature of the vessel axis [3, 4] and on direction changes along the vessel [1]. All measures proposed, to our best knowledge, represent blood vessels as one-dimensional curves.

However, the perceived tortuosity of blood vessels seems to depend on thickness as well: in Figure 1(a), for instance, one would attribute a higher tortuosity level to the thicker vessels. Notice that vessel thickness affects also the maximum allowable curvature of the skeleton, which is inversely proportional to thickness (Figure 1(b)). In addition, a curve representation does not capture important information related to boundaries (e.g., aneurysms), which is of course captured by thickness analysis (Figure 1(c)).

This paper presents a new tortuosity measure using skeleton curvature *as well as* vessel thickness, as opposed to currently used measures [2–4]. We also propose an algorithm identifying automatically a vasculature segment connecting two points specified manually. The algorithm computes a skeletal (medial axis) representation of the image on which all terminal and branching points are located. This is then converted to a graph representation including connectivity information for all vessels. Target segments for tortuosity estimation are identified automatically given two end points, using a shortest-path algorithm.

The remainder of the paper is organized as follows. Section II sketches the algorithm locating a blood vessel segment between two points selected manually, and Section III gives details. Section IV introduces our tortuosity measure and recalls the standard ones. Section V presents results and compares the tortuosity estimates generated by our system and by a clinician.

2 OUTLINE OF THE ALGORITHM

The algorithm is organized in two stages: (a) location and characterization of a target vessel (Section III), and (b) estimation of vessel tortuosity using the novel metric proposed (Section IV).

*Email:ha19@hw.ac.uk, e.trucco@hw.ac.uk, Bal.Dhillon@luht.scot.nhs.uk, T.J.MacGillivray@ed.ac.uk, s0092761@sms.ed.ac.uk.

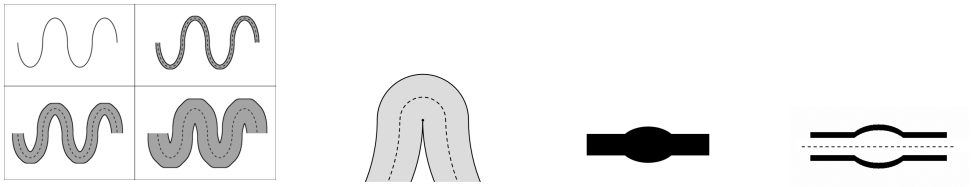


Figure 1. (a) (b) (c)
(a):The larger the thickness, the higher the tortuosity. **(b)**:Thickness limits skeleton bending, i.e., the max curvature achievable. **(c)**:Contour-related features like aneurysms are not captured by skeleton analysis.

The input is a binary image of the retinal fundus, highlighting the vascular network. A region of interest (ROI) containing the vessels for which tortuosity is to be estimated is selected manually using a GUI. All subsequent operations take place within the ROI.

We first perform binary skeletonisation [5] and compute a graph representation of the vessel network. Crucially, local thickness information are preserved for all arcs using the Euclidean distance transform of the binary image within the ROI. The Terminal and branching nodes are then located. Connectivity information obtained so far is organized into an adjacency matrix. The endpoints of a target vessel, for which tortuosity must be estimated, are then selected by the user via the GUI. Dijkstra's algorithm is applied to find the shortest path between the selected endpoints. The path gives the vessel segment on which tortuosity is estimated. Thickness information is used to compute the local curvatures along the two borders of the candidate vessel. The proposed tortuosity metric is an avereged sum of the computed borders curvatues.

3 LOCATION AND CHARACTERIZATION OF TARGET VESSELS

3.1 Skeletonisation

The input image is a binary map of the retinal vasculature. We have used the publicly available DRIVE set by Staal et al. [6]. Figure 2(b) shows an example of binary map for the fundus image shown in 2(a). Skeletonisation is performed by the MATLAB *bwmorph* function, based on iterative deletion of pixels and preserving 8-neighbor connectivity. The result for the input image above is shown in Figure 2(c).

3.2 Branching and terminal nodes detection

Using the skeleton image, terminal and branching points are detected. We count the number *t* of transitions from black to white moving clockwise around the 8-neighborhood of a point, which is classified as follows:
t=1: determines a terminal node;*t*=0, 2: determines a non-significant point;$t \geq 3$: determines a branching point.
Results for the example above are shown in Figure 2(d).

3.3 Building the adjacency matrix

We now have the skeleton structure of the vasculature inside the ROI, with arcs associated to portions of detected vessels. Vertex connectivity, i.e., the existence of an arc linking two vertexes, is checked by deleting the vertexes and their 8-neighborhood, then labelling the connected components in the resulting image L; these are the edges of the graph, and a different label is given to each of them. Two vertexes i and j are connected if they lie at the 2 extremities of a constant-label component of L.

We number all the vertexes of the graph and build a connectivity matrix, M_C, in which: $M_C(i,j) = 1$ if vertexes i and j are connected, i.e., belong to the same arc, and $i \neq j$;
$M_C(i,j) = \infty$ if i and j are not connected;
$M_C(i,j) = 0$ if $i = j$;
$M_C(i,j) = M_C(j,i)$, as connectivity is symmetric.

3.4 Selection of blood vessel segments

A blood vessel segment on which to estimate tortuosity is chosen by selecting two endpoints manually. If there is no branching node between the endpoints, the target segment is given simply by the graph arc containing the two points. Otherwise the segment is given by the shortest path between the endpoints, estimated by Dijkstra's algorithm using the connectivity matrix M_C. Figure 2(e) shows an example superimposed to the fundus image in Figure 2(a).

Figure 2. (a) (b) (c) (d) (e)
(a): fundus image (DRIVE set). **(b)**: binary image (segmented vasculature). **(c)**: extracted skeleton. **(d)**: skeleton with branching and endpoints highlighted. **(e)**: vessel segment between manually selected endpoints.

4 ESTIMATION OF VESSEL TORTUOSITY

4.1 Boundary localization and curvature estimation

Computing the proposed tortuosity measure for the target segment requires estimates of skeleton curvature and local vessel width.

The set of coordinates of the skeleton pixels on the shortest path, $[X_s(j), Y_s(j)]$, is output by the previous stage and used to estimate curvature; j is a parameter identifying position on the curve, To achieve sensible estimates from the numerical curve, the skeleton is approximated locally by a spline. We used a sampling distance of 3 pixels, which achieves an adequate estimation of tortuosity and limits the effect of sampling noise.

The boundary pixels corresponding to each skeleton point are computed as follows. Let R_s be the vector of the radius (i.e., half-width) values of the vessel cross-section at $[X_s(j), Y_s(j)]$. The corresponding boundary points, $[X_{bi}(j), Y_{bi}(j)]$, $i = 1, 2$, are defined as the intersection of the line normal to the skeleton at j and the contours of the vessel in the binary vessel image. Their position is given by (omitting j for simplicity) $X_{bi} = X_s \mp \frac{R_s Y'_s}{\sqrt{X'^2_s + Y'^2_s}}$ and $Y_{bi} = Y_s \pm \frac{R_s X'_s}{\sqrt{X'^2_s + Y'^2_s}}$ where the primes indicates differentiation with respect to the curve parameter, j.

4.2 Comparing tortuosity measures

4.2.1 Tortuosity from skeleton only

We use a general measure of tortuosity SCM_p based on local curvature [3], the p-root mean value of curvatures $C_s(j)$. This measure is based only on the skeleton points, $[X_s, Y_s]$.
$C_s(j) = \frac{X'_s(j)Y''_s(j) - X''_s(j)Y'_s(j)}{[X'_s(j)^2 + Y'_s(j)^2]^{3/2}}$ and $SCM_p(V) = SCM_p([X_s, Y_s]) = (\sum_j |Cs(j)|^p)^{\frac{1}{p}}$

Where p is a strictly positive integer; we discuss briefly its value below.

4.2.2 Tortuosity combining thickness and curvature

We define a tortuosity measure that takes into account the skeleton structure as well the vessels thickness. The intuition is that, at a parity of skeleton, the curvature of the vessels boundaries (walls) changes with thickness. Our measure, the Absolute p-Curvature Width Metric (CWM_p), is the sum of the averaged curvatures *of pairs of corresponding boundary points*:

$$CWM_p(V) = CWM_p([X_s, Y_s], R_s) = (\sum_j \frac{|C_{b1}(j)|^p + |C_{b2}(j)|^p}{2})^{\frac{1}{p}}$$

where C_{b1} and C_{b2} are, respectively, the local curvatures of the two vessel boundary points, $b1$ and $b2$.
This metric has the following properties:

1. $CWM_p([X_s, Y_s], 0) = SCM_p([X_s, Y_s])$: *the measure coincides with the sum-of-curvature metric when the vessel width is zero.*
   Proof: Trivial, as the vessel coincides with its skeleton:$CWM_p([X_s, Y_s], 0) = (\sum_j \frac{|C_{b1=s}(j)|^p + |C_{b2=s}(j)|^p}{2})^{\frac{1}{p}}$.

2. $CWM_p([X_s, Y_s], R_s)$ *is an increasing function of* R_s.
   Proof: Without loss of generality we assume a constant radius, $R_s = D$. We first proof that CWM_p increases with D when the skeleton is an arc of a circle of radius $R0$, ($D \leq R0$), then we generalize to any C^2 curve.
   Let δ be the length of a skeleton arc, and $b1$ and $b2$ the inner and outer boundary points. We have:

   $$C_{b1} = \frac{1}{R0-D}, \quad C_{b2} = \frac{1}{R0+D}, \quad C_s = \frac{1}{R0}$$

   Thus, omitting some algebra for brevity, $CWM_p([X_s, Y_s], D) = (\frac{\delta}{2})^{1/p} \frac{R0}{R0^2 - D^2} \sum_{2k=0}^{2k<=p} C_p^{2k} (D/R0)^{2k}$ which is an increasing function of D. This result generalizes to any C^2 curve considering that, locally, the curve is coincident with an arc of its osculating circle, and applying the proof above to the latter.

3. Increasing p puts more emphasis on high-curvature vessel segments.
   Proof: It follows from the definition of $CWM_p(V)$, where p is a strictly positive integer, that high values of $|C_{b1}(j)|and|C_{b2}(j)|$ contribute with a polynomially increasing term to the tortuosity value.

Figure 3(a) plots CWM_p for the vessel skeleton in Figure 2(e), with $p = 1$ and increasing thickness. Notice that CWM_p increases with thickness, as expected (property 2).

5 RESULTS

5.1 Reference standard

Tests were carried out on ten images of the DRIVE set. On each image, five vessels of interest were highlighted using the algorithm in Section II, for a total of 50 vessel segments spanning different widths, lengths and tortuosity levels. Figure 3(b),(c) shows some examples

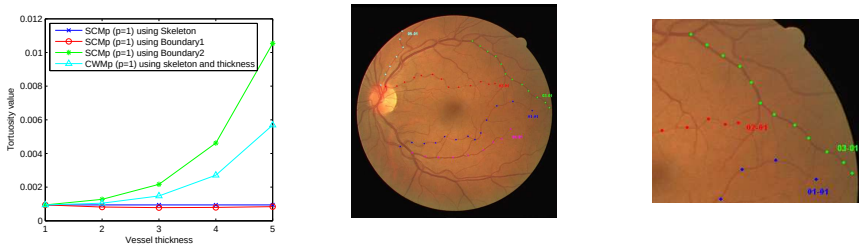


Figure 3. (a) (b) (c)
(a): Plot of SCM_p ($p = 1$) for skeleton and boundaries, and of CWM_p, against increasing thickness (vessel radius) values. The skeleton is that of Figure 2(e). **(b)**: DRIVE image with hilighted vessel segments. **(c)**: Zoom on the same image.

To establish a standard against which to compare automatic results, the tortuosity of each vessel was assessed by the clinical author (BD), a retinal specialist with experience of diseases causing tortuosity in retinal blood vessels. Three levels were used: not tortuous, low tortuosity, high tortuosity.

5.2 Tests and results

Following [3], we used a logit model [7] which computes a weighted function $x = a + bt_v$, where t_v is the tortuosity of a vessel segment and a, b are weights, then applies the logistic function $h(x) = \frac{\exp(x)}{1+\exp(x)}$, $h(x) \in [0, 1]$.
A vessel is assigned to one of the three tortuosity classes depending on h: not tortuous if $h \leq 0.33$, low if $0.33 < h < 0.67$, high if $h \geq 0.67$. The weights a and b were optimized on a subset of the vessels [3].

Table 1 reports percentages of automatic classifications in agreement with clinical judgement. We show results for our measure, with $p = 1$ and $p = 2$, and for the distance metric using only the skeleton [8, 9]. It can be seen that, for this data set, our thickness-dependent measure has an overall better agreement with clinical judgement. Our method achieved a total error rate of 11.2% for $p = 1$ and 7.6% for $p = 2$, giving an overall accuracy of 92.4% for $p = 2$. This compares favorably with the best percentages reported for comparable systems, although of course comparisons in different experimental conditions are only suggestive.

		Clinical Absent	**Ground** Low	**Truthing** High
Automatic	Absent	$80\%, 90\%, 90\%$	$18\%, 8\%, 10\%$	$2\%, 2\%, 0\%$
Classification	Low	$6\%, 2\%, 19\%$	$90\%, 92\%68\%$	$4\%, 6\%, 13\%$
Results	High	$0\%, 0\%, 14\%1$	$6\%, 4\%, 26\%$	$94\%, 96\%, 60\%$

Table 1. Confusion matrix of classifications between automatic method (rows) and clinical grading(columns) using CWM_p for $p = 1$, CWM_p for $p = 2$ and the distance metric, respectively.

6 CONCLUSIONS AND FUTURE WORK

We have presented a new measure for evaluating tortuosity of retinal blood vessels, combining thickness and curvature information. Thickness is considered through its effect on curvatures at vessel boundaries. Initial tests with 50 vessel segments from DRIVE images, of varying thickness, length and tortuosity, suggest that our measure generates results in better agreement with clinical judgement than based only on vessel skeleton.

Future work will seek to corroborate this conclusion with more extensive data sets and against results of various curvature-based measures. We intend also to identify automatically the optimal value of p via an optimization similar to the one used to determine the logit weights.

References

1. E. Bullitt, G. Gerig, S. Pizer et al. "Measuring tortuosity of the intracerebral vasculature from mra images." *IEEE Trans. Med. Img.* **22**, pp. 1163–1171, 2003.
2. C. Heneghan, J. Flynn, M. OKeefe et al. "Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis." *Med. Img. Analys.* **6**, pp. 407–429, 2002.
3. W. Hart, M. Goldbaum, B. Cote et al. "Measurement and classification of retinal vascular tortuosity." *Int. Journ. Medical Informatics* **53**, pp. 239–252, 1999.
4. E. Grisan, M. Foracchia & A. Ruggeri. "A novel method for the automatic evaluation of retinal vessel tortuosity." *Proc. 25th IEEE EMBS* pp. 866–869, 2003.
5. L. Lam, S. W. Lee & C. Suen. "Thinning methodologies-a comprehensive survey." *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**, pp. 879, 1992.
6. J. J. Staal, M. D. Abramoff, M. Niemeijer et al. "Ridge based vessel segmentation in color images of the retina." *IEEE Trans. Medical Imaging* **23**, pp. 501–509, 2004.
7. J. Cramer. "The logit model: An introduction for economics." 1991.
8. N. Patton, T. M. Aslam, T. MacGillivray et al. "Retinal image analysis: concepts, applications and potential." *Prog. Retin. Eye Res.* **25**, pp. 99–127, 2006.
9. M. R. N. Patton, T. MacGillivray, T. M. Aslam et al. "Effect of axial length on retinal vascular network geometry." *Amer. Journ. Ophthalmol.* **140**, pp. 648–53, 2005.

Evaluation of required and available ultrasound image decompression accuracy for decorrelation based distance estimation

Catherine Laporte[a] and Tal Arbel[a*]

[a]Centre for Intelligent Machines, McGill University, Montreal, CANADA

Abstract. In the context of sensorless freehand 3D ultrasound, estimation of out-of-plane transformations between ultrasound images is an important step. This is typically done based on speckle decorrelation algorithms requiring the availability of the envelope intensity signal which has usually been log-compressed by the time it is output by a clinical scanner. This paper studies the effect of decompression errors on the accuracy of decorrelation based distance estimation through simulation experiments. It is found that distance estimation is quite robust to these errors, breaking down when the compression parameter is underestimated by over 30%. Thoroughly testing the current state of the art decompression algorithm shows that despite a tendency to underestimate the compression parameter, the accuracy of this algorithm lies within the tolerance of decorrelation based distance estimation.

1 Introduction

In the last decade, there has been considerable interest in freehand 3D ultrasound systems which estimate the relative out-of-plane motion of ultrasound image frames solely on the basis of image information, without relying a potentially cumbersome external (optical or magnetic) position sensor to track the position of the transducer. This has led to the development of algorithms for estimating the elevational separation (*i.e.* the distance) between ultrasound images based on speckle decorrelation methods [1–4].

Speckle decorrelation methods rely on certain statistical properties of the raw radio-frequency (RF) or echo envelope intensity signal. Unfortunately, these are not the signals supplied by most clinical scanners, which apply non-linear dynamic range compression operations to these signals for visual enhancement purposes. Typically, the exact form of these non-linear operations is not known and must be reverse engineered (inevitably with some error) before applying decorrelation methods. This reverse engineering process is referred to as *decompression*. The aim of this study is to gain an understanding of the relationship between the accuracy of decorrelation based distance estimates and errors in the decompression process.

The intensity of an ultrasound signal backscattered from a given point in space depends on the scatterers located within a certain region around it called a resolution cell. The intensities at neighbouring points in space are correlated due to their overlapping resolution cells. Under conditions known as fully developed speckle, this correlation decreases with the extent of the overlap (*i.e.* distance) in a predictable, transducer dependent manner. A typical implementation of distance estimation by speckle decorrelation involves the construction of a transducer dependent *decorrelation curve* which relates the Pearson correlation coefficient ρ between two ultrasound envelope intensity signals to the spatial distance d between them. Assuming a Gaussian shaped resolution cell, this relationship is well described by a truncated Gaussian curve [1], *i.e.*

$$\rho \approx \exp\left(-\frac{d^2}{2\sigma^2}\right), \tag{1}$$

where σ is related to transducer beamwidth. This curve can be constructed during a calibration step from a set of training images of fully developed speckle by fitting a Gaussian function to average estimates of ρ found for images separated by known elevational distances, as described in [2]. Relative motion between pairs of test image frames is then found by measuring their correlation coefficient ρ, and looking up the corresponding separation distance on the Gaussian decorrelation curve, according to

$$d = \sigma\sqrt{-2\ln\rho}. \tag{2}$$

This methodology operates on the echo intensity envelope signal, which is typically not available from clinical ultrasound scanners. Most clinical scanners apply logarithmic compression to the echo data in order to reduce its the dynamic range for display on a monitor and preserve low contrast boundaries for better visualisation. It has been suggested [5, 6] that the resulting signal is well modeled by a mapping of the form

$$p = D\ln I + G, \tag{3}$$

*Please send correspondence to cathy@cim.mcgill.ca

where p is the value of the pixel displayed on the monitor, I is the envelope intensity signal and parameters D and G are generally hidden from the user of the ultrasound machine. Provided D is known, one can obtain a signal with statistical properties suitable for use with speckle decorrelation algorithms. Thus, it is the value of D which must be reversed engineered during the decompression process before applying speckle decorrelation methods.

This paper studies the relationship between the accuracy of decorrelation based distance estimates and errors in the decompression process. The accuracy required for distance estimation is compared to the accuracy provided by the current state of the art decompression algorithm developed by Prager *et al.* [6] in a variety of conditions. Experiments are presented in the form of a simulation study. It is shown that speckle decorrelation is robust to decompression errors, except when D is underestimated by 30% or more, and that the accuracy of the state of the art decompression algorithm falls within that tolerance, confirming its appropriateness in the context of decorrelation based distance estimation.

The remainder of the paper is organised as follows: section 2 describes the simplified model of ultrasound imaging used to carry out the simulations presented in the paper. Based on simulations, section 3 discusses the effects of decompression errors on the decorrelation based distance estimation. Finally, section 4 presents results from further simulations designed to evaluate the decompression algorithm by Prager *et al.*.

2 Simulation model

In order to measure the influence of decompression errors on the accuracy of decorrelation based distance estimation as well as the accuracy of the chosen decompression algorithm, the experiments presented in this paper use simulated ultrasound data for which ground truth quantities of interest are known with absolute certainty. The simulations are based on a simplified 2D model of ultrasound imaging [7]. The transducer point spread function (PSF) is modeled as a sinusoid modulated by a 2D Gaussian envelope. Throughout this paper, the central frequency is assumed to be 7.5 MHz, and the Gaussian envelope has a standard deviation of 0.2 mm in the direction of sound propagation, and a standard deviation of 0.4 mm in the perpendicular direction. The speed of sound is taken to be 1446 m/s.

A dense set of RF signals is generated by convolving the PSF with a set of 2D delta functions representing point scatterers. The sampling frequency of the RF signals is 66 MHz in this paper. A subset of RF signals (sampled at regular intervals) are then envelope detected using the Hilbert transform and squared to form the intensity signals needed by the speckle decorrelation algorithm (hereafter to be called *intensity vectors*). This is illustrated in figure 1.

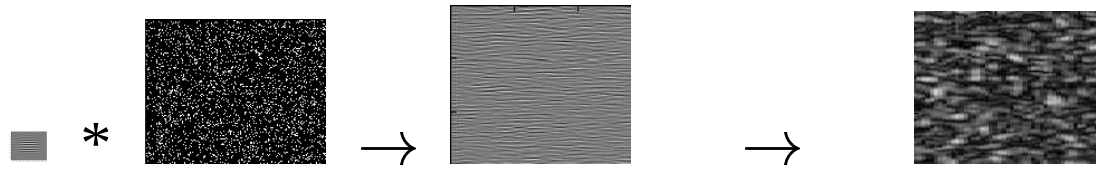


Figure 1. The ultrasound simulation process.

The point scatterers are placed randomly according to Cramblitt and Parker's marked regularity model [8]. The model has a scatterer regularity parameter which is inversely proportional to the variance in scatterer separation. A value of 1 for this parameter corresponds to Rayleigh scattering (*i.e.* fully developed speckle). Scatterers tend to group in clusters as the regularity parameters takes values below 1 (model typical of cardiac and obstetric imaging), while they tend towards periodicity (*i.e.* coherent scattering) for values greater than 1 (model typical of abdominal imaging).

3 The effect of decompression errors on distance estimates

In order to evaluate the effect of decompression errors on the accuracy decorrelation based distance estimation, two ultrasound envelope intensity data sets were simulated with Rayleigh scattering conditions (these are a theoretical requirement for the speckle decorrelation approach, at least, in its basic form). The data sets each consisted of an image made of 1024 regularly spaced intensity vectors of 4096 samples. Neighbouring intensity vectors were separated by a distance of approximately 0.04 mm. Both data sets were compressed with the same value of D, using (3) with a constant offset G. Both data sets were then decompressed using the same value of D with varying amounts of error added, according to $I = \exp(p/(1+\epsilon)D)$, where ϵ is the relative decompression error. The first data set was used to estimate a decorrelation curve. The accuracy of distance estimates based on this decorrelation curve was then evaluated by estimating distances between pairs of intensity vectors in the second data set. The experiment was carried out for

integer values of D between 12 and 20.

The results of this experiment are presented in figure 2. The top chart indicates that for small separations, the distance estimates have very little bias, irrespective of decompression error. The speckle decorrelation algorithm breaks down as theoretically expected for distances larger than one half beamwidth (approximately $0.4 \times 2.35 = 0.94$ mm for these simulations). As separation approaches and grows beyond this limit, a negative bias appears, which is explained by the increasing number of estimates of ρ which fall below 0, causing invalid decorrelation curve lookups. This causes the average error to be dominated by cases where ρ is overestimated (and the distance, underestimated, see (2)) [4]. The bottom of figure 2 shows the average number of invalid lookups as a function of distance for the same errors in D. According to this plot, invalid lookups occur more frequently and for significantly shorter distances when D is underestimated by 30% or more. Overestimating D does not affect the average distance error very much, even for large amounts of error. The middle of figure 2 shows the standard deviation of the error in distance estimates as a function of true distance. The increase of standard deviation with distance occurs at a much faster rate when D is underestimated by 30% or more, while overestimating D even by large amounts has very little effect.

4 Accuracy of the state of the art decompression algorithm

The current state of the art decompression algorithm was developed by Prager *et al.* [6]. This algorithm estimates D from a homogeneous patch of fully developed speckle in an ultrasound image. If the assumption of fully developed speckle holds true, the envelope intensity signal should have the statistics of an exponential distribution [9]. The decompression algorithm finds a value of D that minimises the least squares error between the estimates of normalised fractional order moments of the decompressed data and those of an exponential distribution.

The algorithm was shown by its authors to find very accurate D for independent samples from an exponential distribution. Due to the coherent nature of ultrasound imaging, however, the samples in an ultrasound image are correlated. In images supplied by clinical scanners, some of this correlation (in the axial direction) disappears due to subsampling, but more correlation is introduced in the lateral direction by interpolating values between the intensity vectors for display.

To evaluate the effect of sample correlation on the accuracy of D, 50 ultrasound images consisting of 25 intensity vectors of 700 samples of fully developed speckle were simulated and then subsampled by a factor of 10 in the axial direction and upsampled by a factor of 3 using linear interpolation in the lateral direction, yielding 5250 samples per image. 50 larger images were also produced using 75 intensity vectors of 2200 samples, yielding images of 49500 samples after subsampling and interpolation. These images were compressed with integer values of D between 12 and 20, rounded to the nearest integer and decompressed using the algorithm by Prager *et al.*. The results are given in table 1, compared with those obtained from 50 similarly sized sets of independent samples drawn from an exponential distribution. The results indicate that D tends to be somewhat underestimated when the algorithm is run on correlated speckle data. The bias and variance in the decompression error are reduced when more samples are available.

	mean error in D	standard deviation of error in D
5000 independent samples	$\sim 0\%$	1.83%
50000 independent samples	$\sim 0\%$	0.56%
5250 pixel speckle image	-4.49%	5.52%
49500 pixel speckle image	-1.82%	2.22%

Table 1. Average decompression error for sets of independent samples from an exponential distribution and sets of simulated ultrasound images with correlation. The errors are given in percentage of the correct value of D.

The decompression algorithm assumes that it is operating on data with the statistics of fully developed speckle. While such data can be obtained with confidence by scanning a speckle phantom, identifying patches of fully developed speckle in clinical images can be rather difficult. The extent to which non-Rayleigh scattering conditions influence the accuracy of the decompression algorithm was evaluated through a further experiment. For this purpose, sets of 50 simulated images were generated as in the previous experiment using different values for the regularity parameter in Cramblitt and Parker's model [8] (see section 2), such that the resulting appearance of the images could be confused with fully developed speckle. Simulations were carried out for simulated images with 5250 and 49500 pixels.

The results are presented in figure 3. Slight overestimation of D occurs for highly clustered scatterers, but in general, D is underestimated. The standard deviation of the decompression error appears to be quite independent of the scattering conditions and is mainly affected by the number of samples used to estimate D. If it is assumed that the error in D

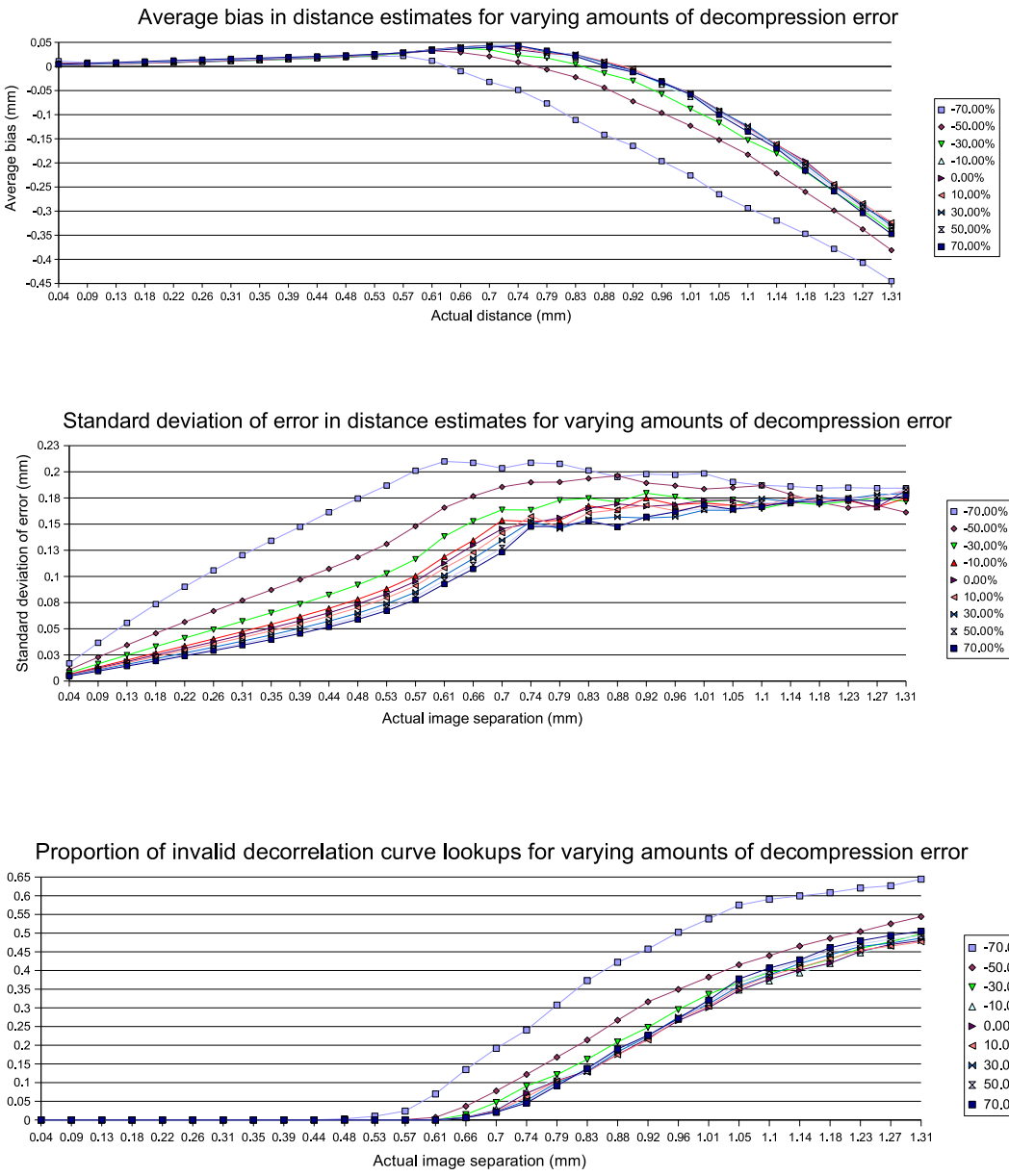


Figure 2. Top: Average error in distance estimates for varying amounts of decompression error. Middle: Standard deviation of error in distance estimates for varying amounts of decompression error. Bottom: Proportion of invalid decorrelation curve lookups for varying amounts of decompression error.

varies within two standard deviations of the mean error, then the error should vary roughly between -20% and +10% if an image patch resembling fully developed speckle with 5250 pixels is used. Most of the time, the decompression error is within the tolerance range for decorrelation based distance estimation. That is, the distance estimates will not differ significantly from those obtained with perfectly accurate decompression.

5 Conclusions

Studying the relationship between the accuracy of decorrelation based distance estimation and the accuracy of the decompression parameter reveals that distance estimation is quite robust to decompression errors, and begins to break down when the decompression parameter D is underestimated by about 30%. Thorough experimental evaluation of the

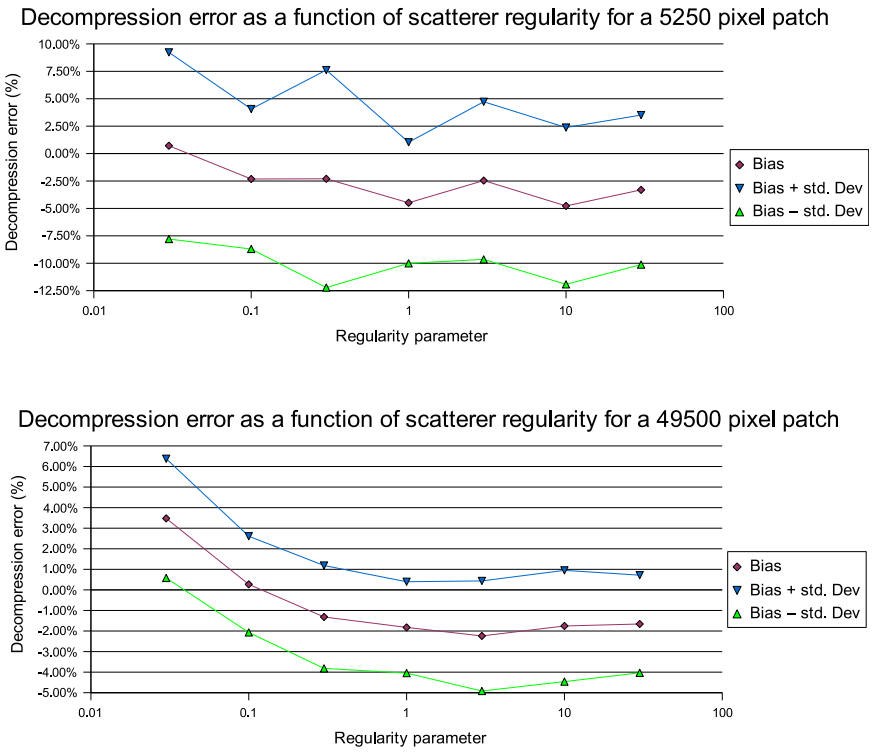


Figure 3. Decompression error as a function of scatterer regularity. The regularity parameter is plotted on a logarithmic scale. Top: for a 5250 pixel image. Bottom: for a 49500 pixel image.

state of the art decompression algorithm in a variety of conditions shows that although it tends to underestimate D, the inaccuracies are within the tolerance of decorrelation based distance estimation. Thus, distance estimates would not be significantly improved by more accurate decompression. This is good news, as it confirms that reasonably accurate results can indeed be obtained without access to signals internal to clinical ultrasound scanners.

Acknowledgements

Catherine Laporte is funded by a doctoral scholarship from the Natural Science and Engineering Council of Canada.

References

1. J.-F. Chen, J. B. Fowlkes, P. L. Carson et al. "Determination of scan-plane motion using speckle decorrelation: theoretical considerations and initial test." *International Journal of Imaging Systems and Technology* **8(1)**, pp. 38–44, 1997.
2. T. A. Tuthill, J. F. Krücker, J. B. Fowlkes et al. "Automated three-dimensional US frame positioning computed from elevational speckle decorrelation." *Radiology* **209(2)**, pp. 575–582, 1998.
3. R. W. Prager, A. H. Gee, G. Treece et al. "Sensorless freehand 3-D ultrasound using regression of the echo intensity." *Ultrasound in Medicine and Biology* **29(3)**, pp. 437–446, 2003.
4. W. Smith & A. Fenster. "Statistical analysis of decorrelation-based transducer tracking for three-dimensional ultrasound." *Medical Physics* **30(7)**, pp. 1580–1591, 2003.
5. D. Kaplan & Q. Ma. "On the statistical characteristics of log-compressed rayleigh signals: theoretical formulation and experimental results." *Journal of the Acoustical Society of America* **95(3)**, pp. 1396–1400, 1994.
6. R. W. Prager, A. H. Gee, G. M. Treece et al. "Decompression and speckle detection for ultrasound images using the homodyned k-distribution." *Pattern Recognition Letters* **24**, pp. 705–713, 2003.
7. T. A. Tuthill, R. H. Sperry & K. J. Parker. "Deviations from Rayleigh statistics in ultrasonic images." *Ultrasonic Imaging* **10**, pp. 81–89, 1988.
8. R. M. Cramblitt & K. J. Parker. "Generation of non-Rayleigh speckle distributions using marked regularity models." *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* **46(4)**, pp. 867–874, 1999.
9. R. F. Wagner, S. W. Smith, J. M. Sandrik et al. "Statistics of speckle in ultrasound B-scans." *IEEE Transactions on Sonics and Ultrasonics* **30(3)**, pp. 156–163, 1983.

Automatic Framework for Medical Image Registration, Segmentation and Modeling

V Petrović*and T Cootes and C Twining and C Taylor[a]

[a]Imaging Science and Biomedical Engineering, University of Manchester, Oxford Road, Manchester, M13 9PT

Abstract. We describe an framework that simultaneously segments and registers a set of medical images in an automatic manner, incrementally constructing a model of the structure and shape deformations of the set. The framework extends existing groupwise registration and modeling approaches by explicitly modeling the fraction of each tissue type in each voxel, rather than the expected intensity in each voxel. This decouples the model from the effects of the imaging sequence and thus imaging modality. When estimating the optimal deformation field between examples in the set each image is compared to its reconstruction generated from the model segment fractions and the current estimate of its intensity distributions for each tissue type (i.e. an estimate of how the model would appear given the imaging conditions for that image). We also present a method to determine the optimal number of tissue types and fully automate the approach as well as model construction methods that ensure efficient convergeance. We describe the algorithm in detail and present results of applying it to a set of MR images of the brain.

1 Introduction

This paper proposes a fully automatic approach for analysing, understanding and representing structure in groups of medical images. Ideally, given a set of images of different examples of a structure, one would like to derive in an efficient and robust (converging) manner:

- a set of deformation fields defining dense correspondences between the images (registration)
- a classification of the voxels in each image into different tissue types (segmentation)
- a statistical representation of the variability of shape and appearance across the image set (modeling)

There has already been considerable research into techniques that could provide each of these requirements independently. Groupwise non-rigid image registration methods derive a dense, spatial correspondence across sets of images [1–3], segmentation of the brain into tissue types in MR images has been considered in [4, 5], while Statistical Appearance Models [1] capture and describe the appearance (shape and texture) variation of the modelled structure.

A variety of methods have been proposed combining pairs of techniques such as segmentation and registration with active countours [6] and maximum a posteriori segmentation using hidden markov random fields and B-spline non-rigid registration [7]. Models of deformation have been constructed from correspondences estimated by non-rigid registration [8, 9], but it was also shown that by integrating modeling and registration more tightly, by performing them in parallel [1, 10] quantitatively better models result.

In this paper we describe an algorithm which combines segmentation, registration and model building in a single iterative framework that incrementally improves the structure analysis to satisfy all the requirements identified above. The method is described in detail in Section 2 while results of applying it to a set of MR images of the brain are shown in Section 3.

2 Method

An overview of the algorithm is illustrated in Figure 1. A set of N images T_i, $i = 1...N$, (the training set) is assumed to have been roughly aligned (either the imaging protocol is sufficient, or for instance by affine registration to one example using a Mutual Information metric). It is further assumed that the structures in the images consist of M distinct tissue types whose intensity probability density functions have parameters θ_i. In the following, $W_i()$ represents a spatial deformation defining correspondence from a reference frame to the image T_i. The segmentation, registration and modeling proceeds sequentially through the following steps (but see below for more details):

1. Warp each training image T_i into the reference frame using the current estimate of the deformation field. $T'_i = W_i^{-1}(T_i)$.

* v.petrovic@manchester.ac.uk

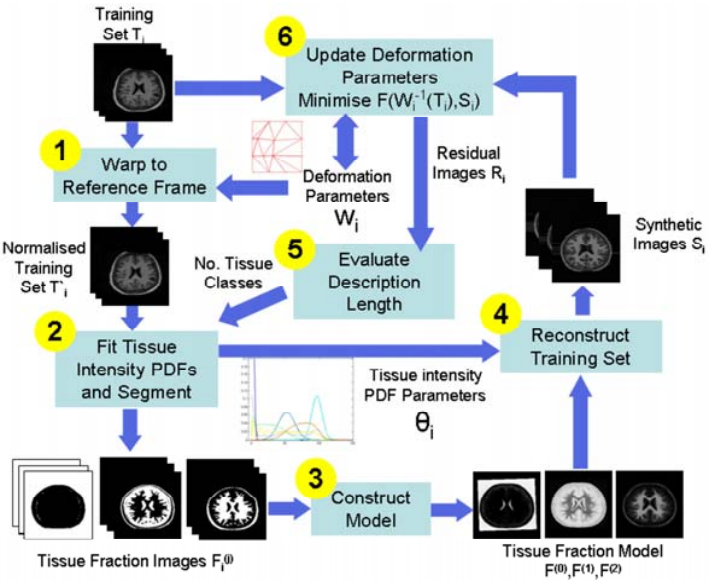


Figure 1. Outline of the algorithm.

2. Compute the intensity histogram of T_i', and fit a mixture model to estimate the distributions and proportions of each pure tissue class (means, SDs and weights encoded in $\theta_i = \{\mu_{ij}, \sigma_{ij}, w_{ik}\}$). Use the resulting distributions to estimate the most probable fraction of each tissue in each voxel, encoded in a set of tissue fraction images $F_i^{(j)}$, $j = 1...M$ for each training example.
3. Combine the tissue fraction images from all examples to construct a model, $\{\hat{F}^{(1)}...\hat{F}^{(M)}\}$.
4. Synthesize a reconstruction of each training set image using the current estimates of the mean pure tissue class distributions, μ_{ij}, and the current tissue fraction model, $S_i = \sum_{j=1}^{M} \mu_{ij}\hat{F}^{(j)}$
5. Find optimal number of tissue classes M by evaluating the Description Length (DL) of the modeled training set representation as a function of M, see [1], and using the one that results in minimum DL (MDL).
6. Update the current estimate of W_i to best register S_i onto T_i, minimising a suitable similarity measure, $D_{im}(T_i, W_i(S_i))$.

During the first pass of the algorithm we apply steps 1 and 2 to all images, then combine them (step 3) into an initial model of the tissue fractions at each voxel in the reference frame. Steps 2 to 5 are then repeated iteratively until an optimum M is found. We then apply steps 1 to 6 (exc. 5) to each image in turn multiple times, performing registration of coarse details in the early stages and progressively finer details as the iterations progresses. The process of shape model building is excluded from the Figure 1 for the sake of clarity, but is an integral part of step 3 and can further be used to constrain the deformation field optimisation in step 6 [10].

2.1 Segmentation

We follow Pokric et al. [4] in assuming that the objects in MR images of the brain are constructed from a small number (M) of different tissue types, and that each voxel contains either a pure tissue or a mixture of at most two different tissues (partial volumes). If we know the distributions of intensities for these classes, we can construct the distribution for a particular fractional distribution by convolution. In the experiments described below we assume that the pure tissues distributions are Normals, $p_i(g) = N(g : \mu_i, \sigma_i^2)$, then the distribution for a partial volume with fraction f of tissue type i and $1 - f$ of type j is given by

$$p_{ij}(g|f) = N(g : f\mu_i + (1-f)\mu_j, f\sigma_i^2 + (1-f)\sigma_j^2) \tag{1}$$

The distribution over all partial volumes containing i and j is given by (2) where we assume all values of f in the range $[0, 1]$ are equally likely ($p(f) = 1$).

$$p_{ij}(g) = \int_{f=0}^{f=1} p_{ij}(g|f)p(f)df = \int_{f=0}^{f=1} p_{ij}(g|f)df \tag{2}$$

If we assume that any voxel contains at most 2 different tissue types, then we need only consider M pure tissue classes with distributions $p_k(g)$, $k = 1..M$, and $M(M-1)/2$ partial tissue classes (enumerated $p_k(g)$,$k = (M+1)..M_t = M(M+1)/2$). Thus the measured image intensity distribution, $h(g)$, can be approximated as a weighted sum where $\theta = \{\mu_i, \sigma_i, w_k\}$ ($i = 1..M, k = 1..M_t$):

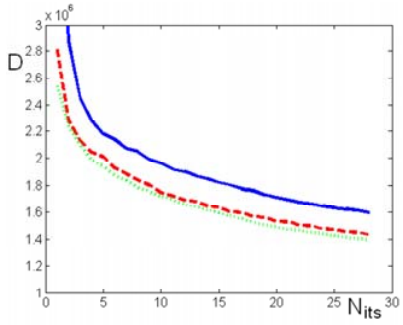


Figure 2. Cummulative obj. function D vs. registration iteration for simple mean (solid), half trimmed mean (dashed) and median model estimation (dotted line)

$$p(g:\theta) = \sum_{i=1}^{M_t} w_i p_i(g) \tag{3}$$

We thus perform an optimisation to estimate the parameters θ which optimise $D_p(p(g:\theta), h(g))$, where $D_p(p,q)$ is a suitable measure of divergence between distributions. The probability that a voxel with intensity g belongs to class k is given by $P_k(g) = w_k p_k(g) / \left(\sum w_k p_k(g)\right)$ (see Figure 4). That voxel can then be classified as belonging to class $k_c = \arg\max_k P_k(g)$. However, we are actually interested in the estimate of the fraction of each pure class tissue (f_i,$i = 1..M$), in the voxel, not the probability of each class. If $k_c \leq M$ then the voxel is a pure tissue, so we define $f_{k_c} = 1$ and $f_{i \neq k_c} = 0$. If $k_c > M$ then the voxel is classified as a partial volume, containing two tissues, say of type i and type j. In this case we wish to find the most likely value of the fractions for each tissue. We define

$$f_i \quad = \quad \arg\max_f p_{ij}(f|g) = \arg\max_f p_{ij}(g|f)p(f)/p(g) = \arg\max_f p_{ij}(g|f) \tag{4}$$

where $p_{ij}(g|f)$ is defined above in Equation 1. We then set $f_j = 1 - f_i$ and $f_{k \neq i,j} = 0$. Figure 4 shows an example of this, demonstrating that tissue probabilities are not the same as estimates of pure tissue fractions. Using this approach we compute M images, $\{F_i^{(1)}, ...F_i^{(M)}\}$, recording the fraction of each tissue type at each voxel in the normalised version of image i (that projected into the reference frame).

2.2 Model Construction

We train a model from the M fractional images from each of the N images. [1] Constraints can be imposed on model construction that would say limit any voxel to a mixture of at most two tissue types but although this directly supports convergence, we found that even the simple mean is powerful enough to drive the process to convergeance.

$$\{\hat{F}^{(1)}...\hat{F}^{(M)}\} = \frac{1}{N}\sum_i \{F_i^{(1)}...F_i^{(M)}\} \tag{5}$$

Greater convergence speed is achieved by using a more robust trimmed mean estimation (evaluated by eliminiting outliers). On a separate data set of 50 synthetic, half light - half dark images that vary randomly in the position intensity tranistion and levels, we found that using only half the samples in mean estimation considerably improves the speed of registration convergence. Convergence for this data set is shown as cumulative objective funcion (D) vs. the registration iteration for the simple mean, half trimmed mean and median on Figure 2.

2.3 Image Reconstruction

In order to align the training set a deformation field is optimised between each T_i and the model (reference frame) embodied in a reconstruction, S_i produced using the model tissue fractions and the estimates of intensity distributions for each tissue type in the current image (i.e. an estimate of how the model would appear given the imaging conditions for that image). Pure tissue types exhibiting Normal distributions are optimally represented by their mean ($\mu_i j$) while partial volume voxels are represented by a sum of pure tissue means weighted by the current model tissue fractions;

$$S_i = \sum_{j=1}^{M} \mu_{ij} \hat{F}^{(j)} \tag{6}$$

For an example, see Figure 5. Ideally S_i is a noise free version of T_i but in practice it starts blurred due to missalignments and gets progressively sharper and closer to T_i as alignment and model improve. Deformation parameters W_i

[1] In practice, when working on image i we construct the model from all $N-1$ other images. Such a leave-one-out approach tends to give more generalisable models and lead to faster convergence.

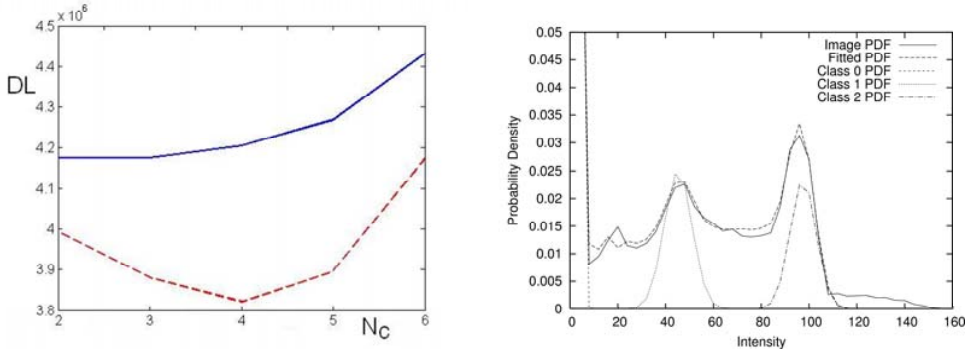


Figure 3. Description Length of modeled training set vs number of tissue classes (N_c) for non-aligned (solid line) and aligned (dashed line) brain images (left); Distribution of intensities for image in Fig. 4 and result of fitting pure and partial tissue class PDFs

are optimised with respect to an objective function measuring similarity between T_i and S_i in the training image frame - $D_{im}(T_i, W_i^{-1}(S_i))$. The deformation is optimised in several stages and at several resolution levels (for the sake of efficiency) by moving individual or groups of control points across the scene, see [10] for details.

2.4 Number of Segment Classes

To initialise the image analysis automaticaly and avoid explicit prior assumptions about image structure we search for an optimal number of tissue types M automatically. For this purpose we use the proposed segmentation-registration algorithm in a Minimum Description Lenght framework, analogously to [1]. We start with the minimum $M = 2$ and perform steps 2 to 4 (Fig. 1 for increasing values of M. For each we evaluate the Decription Lenght of transmiting the training set using the resulting model, given as the cost of the model, model parameters and the residual images ($R_i = S_i - T_i$), $C_{total} = C_{model} + C_{parameters} + C_{residuals}$.

In practice the residual term dominates the cost and improving the model and reconstructed images explicitly improves the Decription Length. If adding another class improves and reconstructed images more than it increases the cost of the model the Description Length decreases. The optimal number of classes is the one that results in a Minimum Description Length (MDL). Note that as we improve the model by aligning images we can realisticaly expect that the optimal value for M might change during registration, see Section 3. For details on how to evaluate C, see [1].

3 Experiments and Results

We applied the method to a set of 32 2D slices of MR brain images (choosing equivalent slices from affine aligned 3D datasets) (Figure 4). We used sum of square differences for both the image similarity, $D_{im}()$, and the divergence between intensity distributions, $D_p()$. Figure 4 shows one of the images from the training set. By fitting pure and partial tissue class distributions to the image histogram (Figure 3), we can estimate the probability that each voxel belongs to each pure tissue class ($P_i(x, y)$ in column 2) and to a partial tissue class ($P_{partial}(x, y)$). We can then estimate the fractions of each tissue type in each voxel ($F_i(x, y)$ in column 3). Finally, combining these fractions with the means of the pure class distributions, we can generate the resconstruction, $S(x, y)$.

Automatic estimation of the number of tissue classes in the data is shown on Figure 3. While the images are not aligned fewer types are sufficient, in this case $M = 3$ is optimal. As the images are brought into alignment, the structures are more coherent accross the examples and further patterns emerge, Figure 3 indicates that four different, consistent tissue types exist in the registered set. As we have not performed any "skull-stripping", this region also contributes to the intensity distributions, in addition to three types (white and grey matter and CSF + background) identified initially.

Figure 5 shows the mean of the model tissue fractions for the three tissue types at the beginning and end of the registration, together with an example of the resulting reconstruction using the means of each pure tissue class. As registration progresses the alignment becomes more accurate, resulting in crisper estimates of the tissue fractions.

4 Discussion

We have demonstrated that it is possible to integrate segmentation, modeling and registration into a single framework and perform complex object structure analysis in a fully automatic manner. By constructing a model of segment

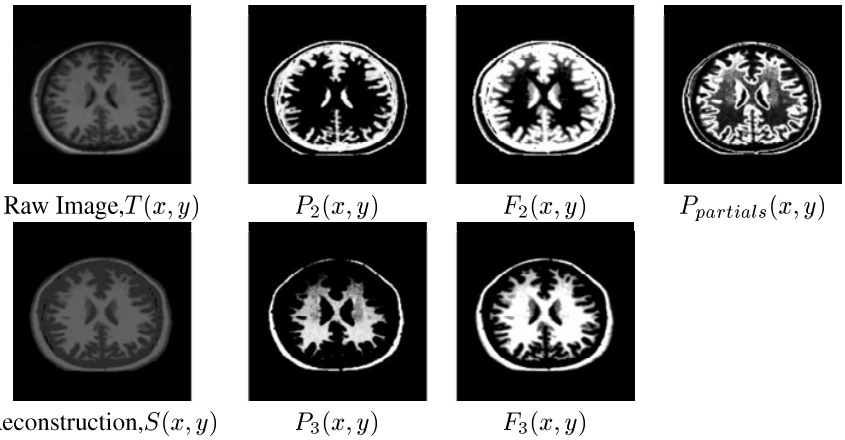


Figure 4. MR brain slice and resulting segmentations. Column 2 shows the pure tissue probabilities, column 3 the estimates of pure tissue fractions.

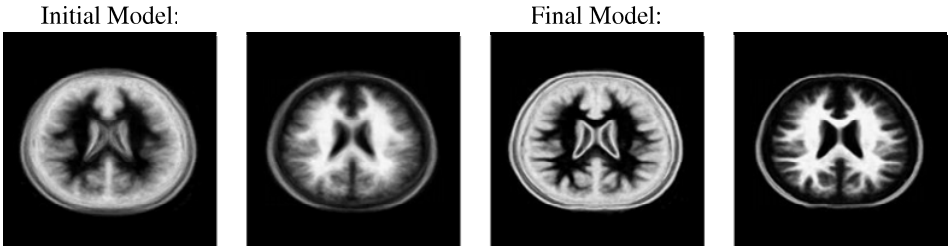


Figure 5. Mean tissue fraction estimates $\hat{F}_2(x,y)$ and $\hat{F}_3(x,y)$ at the beginning and end of registration.

fractions, rather than intensities, we decouple the model from details of the imaging process, and concentrate on explicitly learning tissue structure in medical images. The final model includes both tissue class information and statistical models of shape variation. It should be possible to match such a model to new images taken with different modalities, and we anticipate that such models would have wide application.

We are currently extending the implementation to full 3D data (the extension is natural). Further work will also concentrate on including spatial consistency in estimating the optimal number of tissue types as well as their intensity distributions. Further improvements could also be made by including the current estimates of the mean fraction as a prior when segmenting each image.

References

1. C. J. Twining, T. F. Cootes, S. Marsland et al. "A unified information-theoretic approach to groupwise non-rigid registration and model building." In *Proceedings of Information Processing in Medical Imaging (IPMI)*, volume 3565 of *Lecture Notes in Computer Science*, pp. 1–14. Springer, 2005.
2. K. K. Bhatia, J. V. Hajnal, B. K. Puri et al. "Consistent groupwise non-rigid registration for atlas construction." *Proceedings of the IEEE Symposium on Biomedical Imaging (ISBI)* pp. 908–911, 2004.
3. B. Zitová & J. Flusser. "Image registration methods: A survey." *Image and Vision Computing* **21**, pp. 977 – 1000, 2003.
4. N. A. T. M. Pokric & A. Jackson. "The importance of partial voluming in multi-dimensional medical image segmentation." In *Proceedings of Information Processing in Medical Imaging (IPMI)*, volume 2208 of *Lecture Notes in Computer Science*, pp. 1293–1294. Springer, 2001.
5. Y. Zhang, M. Brady & S. Smith. "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm." *IEEE Transactions on Medical Imaging* **20**, pp. 45 – 57, 2001.
6. A. Yezzi, L. Zollei & T. Kapur. "A variational framework for inregrating segmentation and registration through active contours." *Medical Image Analysis* **7**, pp. 171–185, 2003.
7. D. R. C. Xiaohua, M. Brady. "Simultaneous segmentation and registration for medical image." In *Proceedings of MICCAI 2004*, number 3216 in Lecture Notes in Computer Science, pp. 663 – 670. 2004.
8. A. F. Frangi, D. Rueckert, J. A. Schnabel et al. "Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modelling." *IEEE Transactions on Medical Imaging* **21(9)**, pp. 1151–1166, 2002.
9. D. Rueckert, A. F. Frangi & J. A. Schnabel. "Automatic construction of 3D statistical deformation models using non-rigid registration." *Lecture Notes in Computer Science* **2208**, pp. 77–84, 2001.
10. T. Cootes, C. Twining, V. Petrovic et al. "Groupwise construction of appearance models using piece-wise affine deformations." In *Proceedings of the* 16^{th} *British Machine Vision Conference (BMVC)*, volume 2, pp. 879–888. 2005.

Growing Cell Neural Networks for Fully Automatic Shape Modeling

L. Ferrarini[a*] and H. Olofsen[a] and W. M. Palm[b] and M.A. van Buchem[b] and J.H.C. Reiber[a] and F. Admiraal-Behloul[a]

[a]Division of Image Processing, [b] Department of Radiology
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

Abstract. In this paper, we present a new framework for shape modeling and analysis: we suggest to look at the problem from a pattern recognition point of view, and claim that under this prospective several advantages are achieved. The modeling of a surface with a point distribution model is seen as an unsupervised clustering problem, and tackled by using growing cell structures. The adaptation of a model to new shapes is studied as a classification task, and provides a straightforward solution to the point correspondence problem in active shape modeling. The method is illustrated and tested in 3D synthetic datasets and applied to the modeling of the entire cerebral ventricular system in an elderly population.

1 Introduction

Statistical shape modelling has been increasingly used during the last decade as a basis for image segmentation, interpretation and the studying of shape changes. Many successful 2D-applications have been described in literature (1). The model building requires the establishing of correspondence between shape surfaces over a set of training examples. Defining correspondent points on different shapes is not trivial: in some 2D applications, manual landmark definition might be possible but it becomes unpractical when 3D/4D shapes are considered. Different techniques have been developed to address this problem: some solutions are rooted in computer graphics (2), others in signal/image processing (3), (4), (5), and some have used information theory (6), (7), (8). An important distinction has to be made between *pairwise* and *groupwise* approaches (see (6) and (8)): *groupwise* analysis aims to optimize an objective function over the whole dataset while creating the statistical model, while *pairwise* solutions generally start with a good representative of the dataset and build up a model rooted on it: the method described in this paper is a *pairwise* approach. From a pattern recognition prospective, and considering a pairwise approach, the problem can be summarized in 3 questions: How many points are needed? Where should they be located? How can we define correspondence between points in different models? The question "how many points are needed?" is tackled as an unsupervised learning (clustering) problem where the optimal number of clusters (points) has to be defined. To ensure an optimal location of the points (nodes), the clustering technique needs to be topology-preserving. We consider the point correspondence as a classification problem where the generalization aspect of a classifier is used to match unseen cases to similar previously seen points. We investigated the use of growing artificial neural networks to tackle all three questions. Most of the growing neural networks are variations of the growing cell structure (GCS) introduced by Fritzke in (9). The GCS principle is based on Self Organizing Maps (SOM), which are known to be perfectly topology-preserving: the network preserves neighborhood relations in the data by mapping neighboring inputs onto neighboring nodes in the map. Recently, Marsland et al. presented a self-organizing network that grows when required (SONGWR) (10): SONGWR proved to be (1) more data-driven while growing, and (2) faster in learning input representation, when compared with previous models. In this paper we describe how the SONGWR can be used to create a first point distribution model from a representative instance of the training set (a single case or an average). The adaptation phase of the SONGWR algorithm is then used to deform and match the obtained network to all the instances in the training set. We investigated the robustness and effectiveness of the method in synthetic data sets, and successfully applied it to the modeling of the entire cerebral ventricular system (excluding the fourth ventricle) in magnetic resonance images.

2 Method

The shape modeling technique presented in this work is based on self organizing networks which grow when required (SONGWR), introduced in (10): general aim of these networks is to provide an accurate representation of a given input space. The network is initialized with only few nodes, not connected together, and randomly displaced in the input space. Every time a new data is drawn from the input space, the best matching node is selected among those forming the network at the current state; if the accuracy with which the node represents the data is below a certain threshold a_T, a new node is added, and new connections are created among neighbor nodes.

In case of shape modeling, we start considering a dataset of segmented objects $T = \{O_1, ..., O_n\}$. For each instance in T, the cloud of surface point is detected: $\{P_1, ..., P_n\}$. As a first step, one needs to create a network (or model) of a

*Luca Ferrarini, Division of Image Processing, Leiden University Medical Center. Tel. +31 715263935. Email L.Ferrarini@lumc.nl

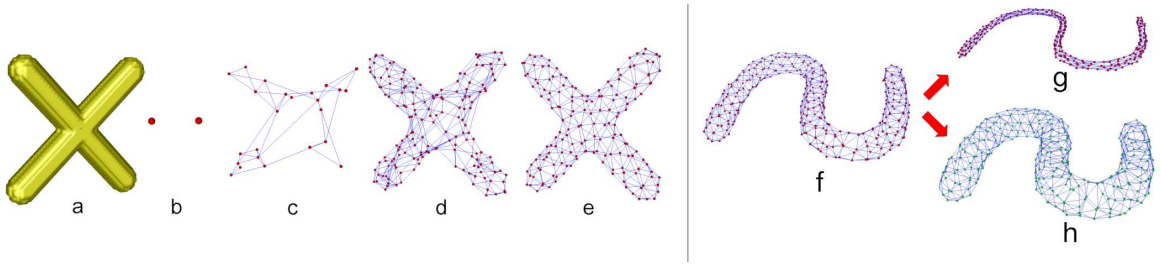


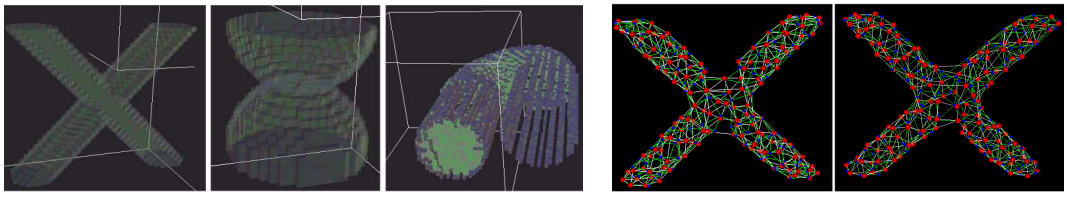
Figure 1. (a)-(e) Unsupervised clustering. (a) Original shape to be modeled. (b)-(e) The neural network grows and adapts to the set of surface points. (f)-(h) Adapting phase: a given model M_i (f) evolves into new comparable models M_j (g) and M_r (h).

given cloud P_i (i.e., an average based on the training set). Our input space is the set of surface points P_i: iteratively, all the surface points are randomly extracted and given to the network. The whole process can be seen as an unsupervised clustering of the surface points: the new nodes added to the network are new clusters, and the new edges encode neighborhood information. The network grows and adapts until convergence is reached: the network structure does not change significantly when new inputs are given. In addition to what suggested in (10), once the convergence has been reached our method undergoes the whole process one more time, using the Mahalanobis distance (instead of the Euclidian one) when looking for the best matching node. We introduced this second step to better represent the surface in locations with high curvature, as suggested in (2). Figure 1.(a)-(e) shows intermediate steps of the unsupervised clustering. After convergence of the Mahalanobis loop, the final model M_i represents the given cloud of points. The original training set includes several instances of similar shapes $T = \{O_1, ..., O_n\}$: in order to perform shape analysis, one needs comparable models of each instance $\{M_1, ..., M_i, ..., M_n\}$. We define two models (M_i, M_j) comparable if each node in model M_i is uniquely associated with a node in M_j, and vice versa. Given the original model M_i and a new cloud of points P_j, a new comparable model M_j is generated through an *adapting phase*: this process can be considered as a classification task, in which the model M_i is the classifier, and the surface points in P_j the objects to classify. For each point, we identify the best matching cluster and adapt it and its neighbors as in SOM. The whole set of surface points is given iteratively until convergence is reached. During the whole process, nodes and edges are not added nor removed: thus, the final model M_j is comparable with M_i. Figure 1.(f)-(h) shows how a given model M_i can adapt to significantly different shapes. For more details on our implementation, the reader is referred to (11).

3 Experiments on synthetic shapes

The evaluation of the algorithm was performed on challenging synthetic shapes: *XShape*, *U-Shaped tube*, and *Chalice shape* (see Fig. 2.a). Experiments were performed to test the stability of the algorithm to random effects and noise, and to assess the ability of the model to adapt to different instances of similar shape capturing real variations.

Both the unsupervised clustering and the adaptation phase introduce a random effect which guarantees a more homogenous growth/adaptation for the model. Reproducibility has to be tested to guarantee similar results in different runs. For each shape, we run the algorithm 20 times using the same cloud of surface points, and evaluated the dissimilarity between models. Table 1 shows the results: the *accuracy* for a model M (given a cloud of points P), and the



(a) (b)

Figure 2. (a) Synthetic datasets: from left to right: *XShape*, *Chalice*, and *U-tube*; (b) Given a *XShape*, several models were built up. For each pair, the dissimilarity was assessed (see eq. 1). The figure on the left shows an overlapping of the two least dissimilar models (diss. = 2.58); figure on the right shows the two most dissimilar models (diss. = 3.59). Even in the worst case scenario (right), the two models (blue and green nodes respectively) are almost indistinguishable.

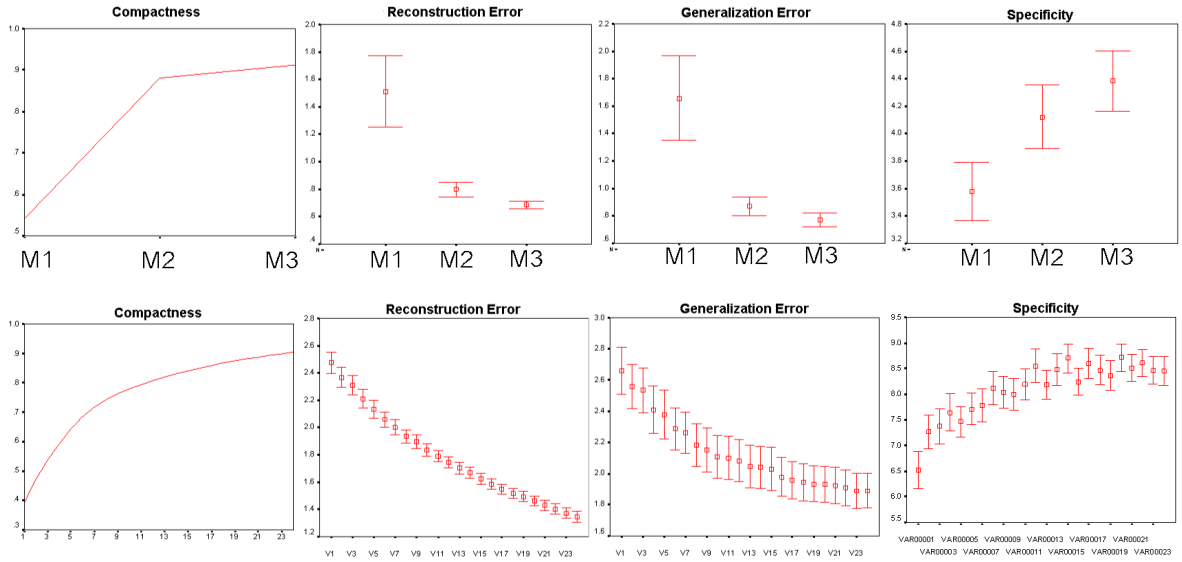


Figure 3. Quantitative analysis for the *Tube-PCA* (first row) and brain ventricles (second row). The errors are given in voxel (first row) and mm (second row). The x axis shows the retained modes of variation. The error bars show the confidential interval at 95%.

dissimilarity between two models $M^{'}$, $M^{''}$ are defined as:

$$Acc(M,P) = \frac{1}{N_{points}} \sum_{k=1..N_{points}} ||p_k - s_k||, \quad s_k = arg\min_{n \in M} ||p_k - n||, \quad p_k \in P, \tag{1}$$

$$d(M^{'}, M^{''}) = Acc(M^{'}, M^{''}) + Acc(M^{''}, M^{'}) + (1 - \frac{E_m}{E_M}), \tag{2}$$

$$E_m = min(E(M^{'}), E(M^{''})), \quad E_M = max(E(M^{'}), E(M^{''})), \quad E(M) = \sum_{q=1..N_{edges}^{model}} ||edge_q||^2,$$

where N_{points} is the number of surface points, s_k is the best matching node in the model given the p_k point of the surface sequence (the Euclidian distance is used), and $E(M)$ is the energy associated with the model (see (10)). Figure 2.b shows a visual example for the dissimilarity between models. The standard deviations for the number of nodes and edges are small compared to the mean value (always below 2% of the average value), and accuracy and dissimilarity are satisfactory considering the total surface. We tested the robustness of our method to noise and missing data. For each shape, starting from the original set of surface points, we built up 20 sequences by applying the following noise: 60% of the points were randomly moved in space (displacement in the range [-2,2] voxels for all the coordinates), and 10% of the nodes were removed. Table 1 shows the results. Introducing noise and missing information in the datasets did not decrease the performances of our method. Finally, we tested the ability of the graph to adapt to different instances of a same class of shapes, capturing the main variations. We created a dataset of 40 tubes, where each tube varies in diameter and length following Gaussian distributions (radius r: $\mu_r = 10, \sigma_r = 2$ voxel, length l: $\mu_l = 125, \sigma_r = 40$ voxel). An average tube was used to train a first model, which was then adapted to the 40 tubes (average accuracy for the adaptation process: $\mu_{acc} = 2.37$, $\sigma_{acc} = 0.2$ voxel). A PCA was performed to analyze the performances of the PDM. The properties we tested are the same presented in (6): *Compactness*, *Reconstruction Error*, *Generalization Error*, and *Specificity Error*. Results are shown in Fig. 3: the PCA highlighted two main modes of variations, corresponding to the two variations artificially introduced in the dataset (changes in length and radius); the third mode of variation reflects noise and errors in registration: its contribution is not significant.

4 Application to a medical dataset

The brain ventricles are located in the center of the brain, and surrounded by gray and white matter structures that are often affected by dementia diseases, as well as by aging: any change of volume or shape occurring in these structures must affect the shape and volume of the ventricles. Studying shape variations within a population might help in understanding how atrophy develops in the brain, either due to dementia or to normal aging. Moreover, the the cerebral ventricular system presents a very challenging shape, highly concave and with strong variability in size (both across subjects, and within different parts of the ventricular system, i.e. lateral ventricles vs. temporal horns). We applied

Table 1. Results for the *reproducibility* and *robustness to noise* tests (*accuracy* and *dissimilarity* in voxel). Results are averaged over 20 runs per shape. Results are provided as mean (standard deviation).

		Reproducibility				Robustness to Noise	
	Surf. Pts.	Avg. Diss.	Accur.	N nodes	N edges	Avg. Diss.	Accur.
X shape	1464	2.84 (0.68)	2.18 (0.005)	146 (0)	372.95 (7.72)	3.09 (0.72)	2.32 (0.01)
Chalice	1208	3.53 (0.94)	2.25 (0.01)	120 (0)	309.65 (4.86)	3.04 (0.71)	2.39 (0.03)
U-Tube	2247	3.33 (0.83)	2.37 (0.007)	223.85 (0.49)	485.25 (10.78)	3.12 (0.74)	2.53 (0.03)

the algorithm on a training set of 67, semi-automatically segmented, brain ventricles (T2-MRI, 91x109x91 voxels at 2x2x2 mm after normalization). After acquiring the clouds of points for each ventricle, we selected a representative one and build up a first model: the optimal number of clusters (nodes) resulted in 755. The model was adapted to the other 66 clouds, and a PCA was performed on the 67 models (average accuracy for the adaptation resulted in $\mu_{acc} = 2.06$, $\sigma_{acc} = 0.22$ mm). Results are reported in Fig. 3, while Fig. 4 shows the first 2 modes of variation. The PCA highlighted 23 modes of variation: when all the modes are considered, the model can successfully generate new instances ($S_p < 9$ mm and $G_e < 2$ mm) and accurately describe the training set ($E_r < 1.4$ mm). The variability found in the population reflects known consequences of atrophy due to aging: the first mode of variation represents a general enlargement of the whole ventricular system, while the second mode of variation highlight elongation/compression of temporal and occipital horns.

5 Discussion

Our method presents several advantages compared with other approaches in literature. In Kaus et al. (2), the authors iteratively reduce the set of original surface points to a subset with dense (sparse) point distributions in areas of high (low) curvature: a triangular mesh is then derived. In our method, all the surface points are used, and the unsupervised clustering detects the areas in which more clusters (nodes) are needed. Differently from (2), we adapt nodes as in SOM: once a node is moved, a parameter controls how many neighbors have to move, and a Gaussian function is used to determine the strength of movements. In Gerig et al. (12), spherical harmonics are used to model a set of surface points. The main advantage is the multi-scale analysis which might highlight deformations not noticeable otherwise; drawbacks of the solution are pre-processing steps needed for *closing* the surface. We have not investigated the multi-scale feature, although we speculate it would be achievable by a reduction in the accuracy for the growing process (a_T). Our solution does not require a *closed* surface. Finally, the parameters used in our approach are the coordinates of the nodes: significant differences in shape are directly associated with positions, (in (12) a visual inspection is needed). In (8), Davies et al. highlight how critical some aspects are for the final model: the number of nodes, their locations, and the correspondence between them strongly influence the final performances. The authors suggest to use *descriptive functions* to identify correspondent points on different shapes, create a first model, and evaluate it through an objective function to be optimized: the search for the best descriptive functions is carried on through genetic algorithms. This approach leads to the distinction between *pairwise* and *groupwise* solutions. *Groupwise* approaches are known to provide better models, optimizing the results on the whole dataset; moreover, they do not require a first representative to start with. *Pairwise* approaches, like the one described in this work, need an initial shape to operate: thus, the final model can be influenced by such a choice. Nevertheless, *pairwise* approaches might overcome some drawbacks related with *groupwise* solutions. The optimization needed by *groupwise* solutions is often achieved by genetic algorithms (see (8)): an optimal result can not be guaranteed and the solution is computationally heavy, a drawback if we aim to develop *learning* systems: how can we improve a given model when a new example becomes available? *Groupwise* solutions

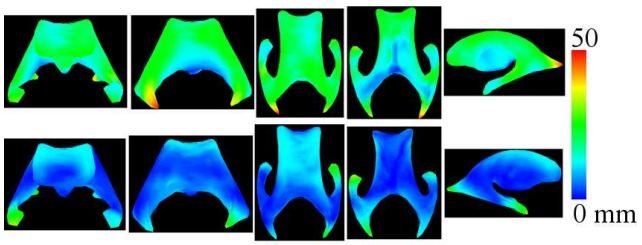


Figure 4. Color-coded maps for the amount of movement of each node. Top row: 1^{st} mode $\pm 3\sigma$. Bottom row: 2^{nd} mode $\pm 3\sigma$. The 1^{st} mode presents a global shrinking/growing behaviour, while the 2^{nd} mode shows elongation/compression of the temporal and occipital horns.

have to undertake the whole optimization process again, with a computation load which grows with the number of shapes. *Pairwise* and *groupwise* solutions should be seen as complementary for many aspects, and the results achieved in one field can improve the other. The results reported in sections 3 and 4 show the stability of our solution and its capability of generating good point distribution models (PDM) for statistical analysis (such as PCA). The stability has been tested on synthetic datasets both with and without noise, while the PDM has been tested both on synthetic dataset and on a medical training set of brain ventricles (lateral and third ventricles included). The PCA could highlight the expected modes of variation for the synthetic data set and known variations in the aging brain ventricles due to brain atrophy.

6 Conclusion

In this paper we presented a new solution, rooted in a pattern recognition framework, for a *pairwise* approach to the shape-modeling problem. The optimal number of nodes needed to describe a shape and their locations are automatically detected. The point correspondent problem is tackled with a self-organizing map approach, which leads to low computational time. We have thoroughly evaluated our method both on synthetic shapes and on a challenging medical application, the entire cerebral ventricular system. Due to its generality, the method can be applied to other anatomical structures. In the future, we will investigate the use of different similarity measurements which could improve the performances in critical applications (i.e., modeling of the cortex surface).

7 Acknowledgment

This work was supported by the Technology Foundation STW (project number LNN. 6122), the technology programme of the Ministry of Economic Affairs, and by Medis medical imaging systems, Leiden, The Netherlands (www.medis.nl).

References

1. T. Cootes, C. Taylor, D. Cooper et al. “Active shape models - their training and application.” *Computer Vision and Image Understanding* **61(1)**, pp. 38–59, 1995.

2. M. Kaus, V. Pekar, C. Lorenz et al. “Automated 3d pdm construction using deformable models.” *Proceedings of the Eigth IEEE International Conference on Computer Vision (ICCV’01)* **1**, pp. 566–572, 2001.

3. A. Pitiot, H. Delingette, A. Toga et al. “Learning object correspondences with the observed transport shape measure.” *Information Processing in Medical Imaging (IPMI)* pp. 25–37, 2003.

4. S. Wrz & K. Rohr. “Localization of anatomical point landmarks in 3d medical images by fitting 3d parametric intensity models.” *Information Processing in Medical Imaging (IPMI)* pp. 76–88, 2003.

5. D. Rueckert, A. Frangi & J. Schnabel. “Automatic construction of 3-d statistical deformation models of the brain using nonrigid resitration.” *IEEE Transactions on Medical Imaging* **22(8)**, pp. 1014–1025, 2003.

6. R. Davies, C. Twining, P. Allen et al. “Shape discrimination in the hippocampus using an mdl model.” *Information Processing in Medical Imaging (IPMI)* **LNCS 2732**, pp. 38–50, 2003.

7. H. Thodberg. “Minimum description length shape and apperance models.” *Information Processing in Medical Imaging (IPMI)* pp. 51–62, 2003.

8. R. Davies, C. Twining, T. Cootes et al. “A minimum description length approach to statistical shape modelling.” *Information Processing in Medical Imaging (IPMI)* **LNCS 2082**, pp. 50–63, 2001.

9. B. Fritzke. “Growing cell structures - a self-organizing network for unsupervised and supervised learning.” *Neural Networks* **7(9)**, pp. 1441–1460, 1994.

10. S. Marsland, J. Shapiro & U. Nehmzow. “A self-organizing network that grows when required.” *Neural Networks* **15**, pp. 1041–1058, 2002.

11. L. Ferrarini, H. Olofsen, M. van Buchem et al. “Fully automatic shape modelling using growing cell neural networks.” *MICCAI* **LNCS 3750**, pp. 451–458, 2005.

12. G. Gerig, M. Styner, D. Jones et al. “Shape analysis of brain ventricles using spharm.” *Published MMBIA, IEEE Computer Society* pp. 171–178, 2001.

Non-Rigid Registration Assessment Without Ground Truth

R. S. Schestowitz[a], C. J. Twining[a], T. F. Cootes[a], V. S. Petrović[a], W. R. Crum[b], and C. J. Taylor[a*]

[a]Imaging Science and Biomedical Engineering, Stopford Building,
University of Manchester, Oxford Road, Manchester M13 9PT, UK.
[b]Centre for Medical Image Computing, Department of Computer Science,
University College London,Gower Street, London WC1E 6BT, UK.

Abstract. We compare two methods for assessing the performance of groupwise non-rigid registration algorithms. The first approach, which has been described previously, utilizes a measure of overlap between ground-truth anatomical labels. The second, which is new, exploits the fact that, given a set of non-rigidly registered images, a generative statistical model of appearance can be constructed. We observe that the quality of this model depends on the quality of the registration, and define measures of model *specificity* and *generalisation* – based on comparing synthetic images sampled from the model, with those in the original image set – that can be used to assess model/registration quality. We show that both approaches detect the loss of registration accuracy as the alignment of a set of correctly registered MR images of the brain is progressively perturbed. We compare the sensitivities of the two approaches and show that, as well as requiring no ground truth, *specificity* provides the most sensitive measure of misregistration. Finally, we use *specificity* and *generalisation* to compare three NRR algorithms.

1 Introduction

Non-rigid registration (NRR) of both pairs and groups of images has been used increasingly in recent years, as a basis for medical image analysis. Applications include structural analysis, atlas matching and change analysis [6]. The problem is highly under-constrained and the plethora of different algorithms that have been proposed generally produce different results for a given set of images [4, 19].

Various methods have been proposed for assessing the results of NRR [9, 11, 13, 16]. Most of these require access to some form of ground truth. One approach involves the construction of artificial test data, which limits application to 'off-line' evaluation. Other methods can be applied directly to real data, but require that anatomical ground truth be provided, typically involving annotation by an expert. This makes validation expensive and prone to subjective error.

We present two methods for assessing the performance of non-rigid registration algorithms; one requires ground truth to be provided, whereas the other does not. We compare the performance of the two approaches by systematically varying the quality of registration of a set of MR images of the brain.

2 Method

The first of the proposed methods for assessing registration quality uses a generalisation of Tanimoto's spatial overlap measure [1]. We start with a manual mark-up of each image, providing an anatomical/tissue label for each voxel, and measure the overlap of corresponding labels following registration. Each label is represented using a binary image, but after warping and interpolation into a common reference frame, based on the results of NRR, we obtain a set of fuzzy label images. These are combined in a generalised overlap score [5]:

$$\mathcal{O} = \frac{\sum_{\text{pairs},k} \sum_{\text{labels},l} \beta_l \sum_{\text{voxels},i} MIN(A_{kli}, B_{kli})}{\sum_{\text{pairs},k} \sum_{\text{labels},l} \beta_l \sum_{\text{voxels},i} MAX(A_{kli}, B_{kli})} \tag{1}$$

where i indexes voxels in the registered images, l indexes the label and k indexes image pairs. A_{kli} and B_{kli} represent voxel label values in a pair of registered images and are in the range [0, 1]. The $MIN()$ and $MAX()$ operators are standard results for the intersection and union of fuzzy sets. The generalised overlap measures the consistency with which each set of labels partitions the image volume. The parameter β_l affects the relative weighting of different labels. With $\beta_l = 1$, label contributions are implicitly volume weighted with respect to one another. We have also considered the cases where β_l weights for the inverse label volume (which makes the relative weighting of different labels equal), where β_l weights for the inverse label volume squared (which gives labels of smaller volume higher weighting) and where β_l weights for a measure of label complexity (which we define arbitrarily as the mean absolute voxel intensity gradient in the label).

*E-mail addresses for correspondence: Roy.Schestowitz@stud.man.ac.uk; Carole.Twining, V.Petrovic, Tim.Cootes, Chris.Taylor@manchester.ac.uk; Bill.crum@cs.ucl.ac.uk

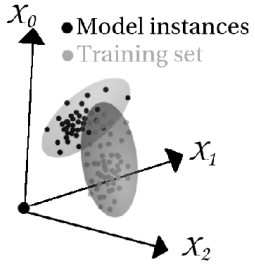


Figure 1. Training set and model in hyperspace

The second method assesses registration in terms of the quality of a generative statistical appearance model, constructed from the registered images – for all the experiments reported here, this was an active appearance model (AAM) [3]. The idea is that a correct registration produces an anatomically meaningful dense correspondence between the images, resulting in a better appearance model of the anatomy. We define model quality using two measures – *generalisation* and *specificity*. Both are measures of overlap between the distribution of original images and a distribution of images sampled from the model, as illustrated in Figure 1. If we use the generative property of the model to synthesise a large set of images, $\{I_\alpha : \alpha = 1, \ldots m\}$, we can define Generalisation G as:

$$G = \frac{1}{n} \sum_{i=1}^{n} \min_{\alpha} |I_i - I_\alpha|, \tag{2}$$

where $|\cdot|$ is a measure of distance between images, I_i is the i^{th} training image, and $\min_\alpha$ is the minimum over α (the set of *synthetic* images). That is, Generalisation is the average distance from each training image to its nearest neighbour in the synthetic image set. A good model exhibits a low value of G, indicating that the model can generate images that cover the full range of appearances present in the original image set. Given a sufficiently large synthetic set, even registered image with differing brightness levels will be paired with a nearby match. Similarly, we can define Specificity S as:

$$S = \frac{1}{m} \sum_{\alpha=1}^{m} \min_{i} |I_i - I_\alpha|. \tag{3}$$

That is, Specificity is the average distance of each synthetic image from its nearest neighbour in the original image set. A good model exhibits a low value of S, indicating that the model only generates synthetic images that are similar to those in the original image set. The uncertainty in estimating G and S can also be computed.

In our experiments we have defined $|\cdot|$ as the shuffle distance between two images, as illustrated for a single pixel in Figure 2. Shuffle distance images are formed by taking the mean of the minimum absolute difference between each pixel/voxel in one image, and the pixels/voxels in a shuffle neighbourhood of radius r around the corresponding pixel/voxel in a second image. When $r \leq 1$, this is equivalent to the mean absolute difference between corresponding pixels/voxels, but for larger values of r the distance increases more smoothly as the misalignment of structures in the two images increases. The effect on the pixel-by-pixel contribution to the shuffle distance image as r is increased is illustrated in Figure 3.

3 Experimental Validation and Application

The overlap-based and model-based approaches were validated and compared, using a dataset consisting of 36 transaxial mid-brain slices, extracted at equivalent levels from a set of T1-weighted 3D MR scans of different subjects. Eight manually annotated anatomical labels were used as the basis for the overlap method: L/R white matter, L/R grey matter, L/R lateral ventricle, and L/R caudate. The images were brought into alignment using an NRR algorithm based on MDL optimisation [18]. The resulting appearance model is shown in Figure 6. A test set of different mis-registrations was then created by applying smooth pseudo-random spatial warps to the registered images. These warps were based on biharmonic Clamped Plate Splines. Each warp was controlled by 25 randomly placed knot-points, each displaced in a random direction by a distance drawn from a Gaussian distribution whose mean controlled the average magnitude of pixel displacement over the whole image. Ten different warp instantiations were generated for each image at each

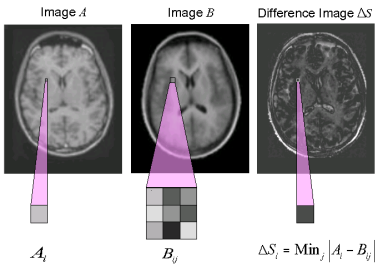


Figure 2. The calculation of a shuffle difference image

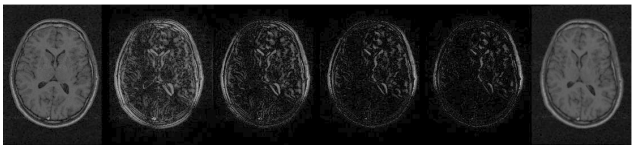


Figure 3. Shuffle difference images between the images on the extreme left and right for $r = 1$ (abs. diff.), 1.5, 2.1 & 3.7 from left to right.

of seven progressively increasing values of average pixel displacement. Registration quality was measured, for each level of registration degradation, using several variants of each of the proposed assessment methods.

To illustrate practical application of the method, we use the brain images described above, and compare the results of three different registration algorithms: a pairwise registration of each training set image to a fixed reference image chosen from the training set, and two MDL groupwise methods, as described above, one with no explicit constraints on the spatial deformations during the registration process (Groupwise 1) and a second which uses a statistical shape model to constrain the allowed spatial deformations between the images during registration (Groupwise 2) [18].

4 Results

The results of the validation experiment are shown in Figure 4. Note that $\mathcal{O}$ is expected to decrease with increasing perturbation of the registration, whilst G and S are expected to increase. All three metrics are generally well-behaved and show a monotonic response to increasing perturbation. This validates the model-based measures of registration quality, which are shown both to change monotonically with increasing perturbation of the registration and to correlate with the gold-standard approach based on manually annotated ground truth.

The results for different values of r (shuffle radius) and β_l all demonstrate monotonic behaviour with increasing perturbation, but the slopes and errors vary systematically. This affects the size of perturbation that can be detected. To make a quantitative comparison of the different methods, we define the sensitivity, as a function of perturbation $(\frac{1}{\overline{\sigma}})\frac{M-M_0}{d}$, where M is the quality measured for a given degree of deformation d, M_0 is the measured quality at registration (no deformation) and $\overline{\sigma}$ is the mean error in the estimate of M over the range.

Sensitivities of the different methods, averaged over the range of perturbations shown in Figure 4, are summarised in Figure 5 for all the methods of assessment. Since sensitivity across the whole range is desired, this average shows that the Specificity measure with shuffle radius 1.5 or 2.1 is the most sensitive of the measures studied, and that this advantage is statistically significant. Exceeding this shuffle radius may lead to performance degradation, as deformations will be obscured by the shuffling.

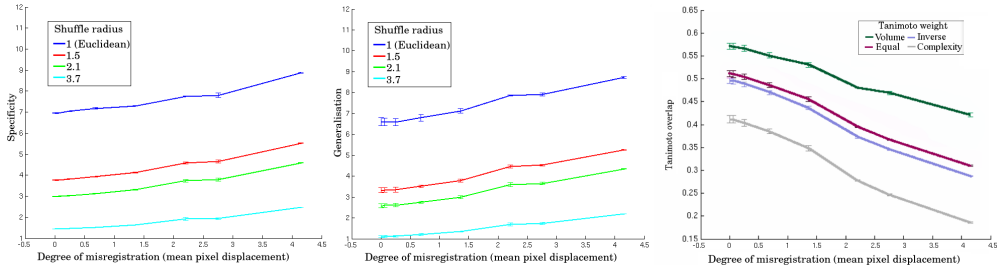


Figure 4. From Left: Specificity (S), Generalisation (G) & Tanimoto overlap ($\mathcal{O}$) as a function of misregistration.

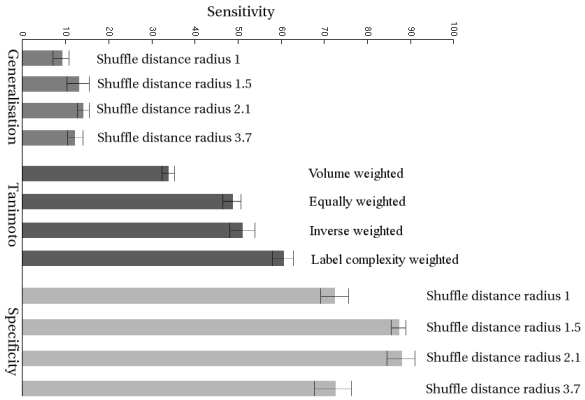


Figure 5. The sensitivities of the different registration assessment methods and their standard errors.

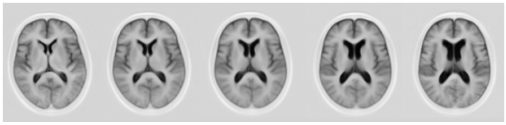


Figure 6. Appearance model constructed from groupwise registered images. First mode of variation is shown, ± 2.5 standard deviations.

The results for different registration algorithms are shown in Figure 7. The specificity obtained for the two groupwise methods is significantly better than that obtained using the pairwise approach, implying better registration, but it is not possible to distinguish between the two groupwise methods. By applying the same NRR algorithms to an annotated dataset, it becomes evident that generalised overlap measures agree with this assessment, for all possible assignments to β_l. As might be expected from the sensitivity results presented above, it is not possible to distinguish between any of the methods using generalisation.

5 Conclusions

We have introduced a model-based approach to assessing the accuracy of non-rigid registration, without the need for ground-truth. We have also described validation experiments where we progressively perturbed the initially good registration of a set of images, and found a monotonic relationship between our model-based measures and the degree of perturbation. We found that this behaviour was qualitatively identical to that obtained using a 'gold standard' method of assessment, based on the overlap of ground-truth anatomical labels associated with the images. A quantitative comparison of the two approaches demonstrated that one of the model-based measures, *specificity*, provides a more sensitive measure of misregistration than the overlap-based approach. This is not as surprising as it might seem at first sight, since the model-based approach uses the full intensity information in the registered images, whereas the overlap-based approach uses a more impoverished representation of image structure. We tested different variants of the two approaches, and found that the model-based approach worked best when shuffle distance was used to measure

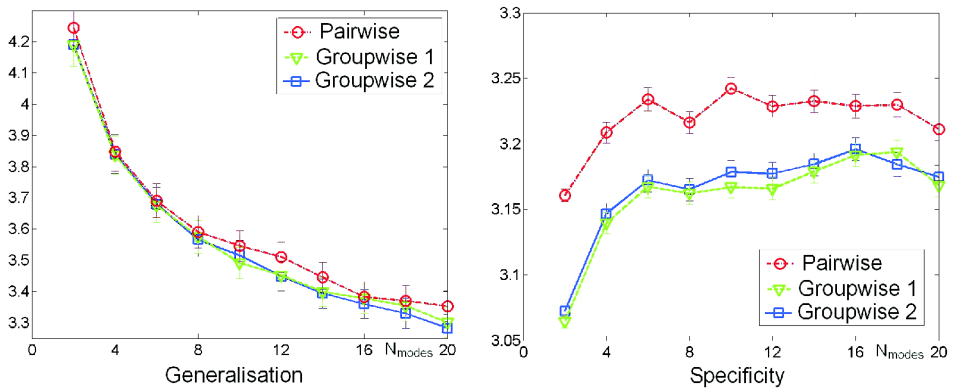


Figure 7. Generalisation and Specificity of the three registration methods as a function of the number of modes included in the appearance model.

separation in image space, whilst the overlap-based approach worked best when a label complexity weighting was applied. We also applied the model based approach to compare three different registration methods, and showed that a groupwise registration approach gave better results than a pairwise approach.

These results are important, because they suggest that the performance of NRR algorithms can be compared objectively, using just the registered images they produce, and that the quality of registration can be assessed in routine applications of NRR, without the need for any additional information. It is important to note that our approach does not depend on the specifics of the registration method used, or on the particular form of generative model constructed from the registered data. It can be applied to the results of registration, whatever the NRR algorithm used, and different forms of generative model could easily be substituted. We demonstrated this in recent experiments, which were omitted due to the limited scope of this paper.

Acknowledgements

The authors would like to thank David Kennedy of the Centre for Morphometric Analysis at MGH for segmented brain data. The work was supported by the EPSRC/MRC-funded Medical Image and Signal IRC (GR/N14248/01), Integrated Brain Image Modelling (EPSRC GR/S82503/01) and Modelling, Understanding and Predicting Structural Brain Change (EPSRC GR/S48844/01).

References

1. M. Beauchemin and K. P. B. Thomson. The evaluation of segmentation results and the overlapping area matrix. *International Journal of Remote Sensing*, 18(18):3895-3899, 1997.
2. T. F. Cootes, C. Beeston, G. J. Edwards, and C. J. Taylor. A unified framework for atlas matching using active appearance models. In *Information Processing in Medical Imaging*, 1613:322-333, 1999.
3. T.F. Cootes, G.J. Edwards and C.J.Taylor. Active appearance models. In *European Conference on Computer Vision*, 2:484-498, 1998.
4. T. F. Cootes, S. Marsland, C. J. Twining, K. Smith, and C. J. Taylor. Groupwise diffeomorphic non-rigid registration for automatic model building. In *European Conference on Computer Vision*, 2034:316-27, 2004.
5. W. R. Crum, O. Camara, D. Rueckert, K. Bhatia, M. Jenkinson, and D. L. G. Hill. Generalised overlap measures for assessment of pairwise and groupwise image registration and segmentation. In *Proceedings of MICCAI*, 3749:99-106, 2005.
6. W. R. Crum, T. Hartkens, and D. L. G. Hill. Non-rigid image registration: theory and practice. *British Journal of Radiology*, 77:140-153, 2004.
7. R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21(5):525-537, 2002.
8. G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using active appearance models. In *European Conference on Computer Vision*, 2:581-595, 1998.
9. J. M. Fitzpatrick and J. B. West. The distribution of target registration error in rigid-body point-based registration. *IEEE Transaction Medical Imaging*, 20:917-27, 2001.
10. A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modelling. *IEEE Transactions on Medical Imaging*, 21:1151-66, 2002.
11. P. Hellier, C. Barillot, I. Corouge, B. Giraud, G. Le Goualher, L. Collins, A. Evans, G. Malandain, and N. Ayache. Retrospective evaluation of inter-subject brain registration. In *Medical Image Computing and Computer-Assisted Intervention*, 2208:258-265, 2001.
12. K. N. Kutulakos. Approximate N-view stereo. In *European Conference on Computer Vision*, 1:67-83, 2000.
13. P. Rogelj, S. Kovacic, and J. C. Gee. Validation of a nonrigid registration algorithm for multimodal data. *Medical Imaging*, volume 4684, 2002.
14. D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8)1014-1025, 2003.
15. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, D. J. Hawkes. Non-rigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712-721, 1999.
16. J. A. Schnabel, C. Tanner, A. Castellano-Smith, M. O. Leach, C. Hayes, A. Degenhard, R Hose, D. L. G. Hill, and D. J. Hawkes. Validation of non-rigid registration using finite element methods. In *Information Processing in Medical Imaging*, 2082:344-357, 2001.
17. M. B. Stegmann, B. K. Ersboll, and R. Larsen. FAME - a flexible appearance modeling environment. *IEEE Transactions on Medical Imaging*, 22(10):1319-1331, 2003.
18. C. J. Twining, T.F. Cootes, S. Marsland, S. V. Petrovic, R. S. Schestowitz, and C. J. Taylor. A unified information-theoretic approach to groupwise non-rigid registration and model building. In *Information Processing in Medical Imaging*, 3565:1-14, 2005.
19. B. Zitova and J. Flusser. Image registration methods: a survey. *Image Vision Computing*, 21:977-1000, 2003.

Image reconstruction with basis functions: Application to real-time radial cardiac MRI

Iason Kastanis,* Simon R Arridge and Derek L Hill

Centre for Medical Image Computing, Department of Computer Science, UCL

Abstract. The imaging of dynamic objects is of high clinical value for diagnosis. In the case of cardiac MRI high temporal resolution cannot be obtained in real-time. Standard methods typically involve long breath-holds and make the assumption that the cardiac motion is perfectly periodic. Due to this fact, there are many clinical situations in which standard methods cannot be applied. The proposed method makes no assumption of this kind and uses a minimal temporal window for the collection of data. Images are reconstructed with the use of local basis functions and the problem of reconstruction is formulated as a system of linear equations. In this paper we present an application of the method for the reconstruction of images in radially sampled cardiac MRI.

1 Introduction

MR data acquisitions are time consuming, because of the need to sample all of k-space in order to reconstruct a fully sampled image. In the case of dynamic imaging, this need is restricting the temporal resolution. Sampling moving objects at many time points is of vital importance to the analysis of these image sequences. In the case of cardiac MRI, the majority of methods uses ECG gating to time-average Fourier data. These methods take advantage of the periodic nature of cardiac motion. Thus patients are required to stay in the scanner for a prolonged period of time performing long breath holds. This limitation excludes people who cannot hold their breath long enough. Free-breathing imaging could potentially be more interesting to clinicians. Fast real-time imaging would also make it possible to image patients with an irregular heart beat.

We propose a method for the reconstruction of images in cardiac MRI that does not assume that the heart beats periodically and does not require long breath-holds. Each image is reconstructed from data acquired within a very small temporal window to reduce any motion artifacts. The proposed method reconstructs images from tomographic data by solving a large system of linear equations. A variety of methods has been proposed based on iterative techniques for the reconstruction of incomplete tomographic data. In [1] a Bayesian inversion method incorporating priors is applied to dental radiology. Level set methods were presented in [2] for 3D shape reconstructions and in [3] for positron emission tomography (PET).

2 Theory

2.1 Model parametrization

If the intensity variation across pixels is considered to be sufficiently smooth, an image can be represented using local basis functions on a regular grid:

$$B_j(\mathbf{r}) = B_0(\mathbf{r} + \mathbf{r}_j) = B_0(\mathbf{r}) * \delta(\mathbf{r} - \mathbf{r}_j)\,, \mathbf{r} \in \Re^2 \tag{1}$$

where r_j is a set of grid points in the pixel domain and B_0 is the central basis function based on the Kaiser Bessel window function [4].

$$B(\mathbf{r}) = \frac{1}{I_m(\alpha)}(\sqrt{1-(d/a)^2})^m I_m(\alpha\sqrt{1-(d/a)^2})\,, \tag{2}$$

I_m is the modified Bessel Function of the first kind, m is the degree, a is the support, d is the distance from the centre of the basis function and α is a shape parameter. The choice of the optimal combination of the support a and the shape parameter α can be found in [5]. Analysis of the Kaiser-Bessel functions and their properties can be found in [6]. Another option for the basis function B are the Wendland functions ([7] and [8]) and a computationally faster alternative are the linear and Gauss radial basis functions [5].

The object f can then be approximated as:

$$\tilde{f} = \sum_{k=1}^{N_p} B_k(\mathbf{r}) p_k \tag{3}$$

where $p_k \in P$ are weights of the basis functions.

*email: j.kastanis@cs.ucl.ac.uk

2.2 Linear least squares image reconstruction

The data g can be expressed as:

$$g = \mathcal{F}f \tag{4}$$

where the operator $\mathcal{F}$ is a radially undersampled 2D Fourier transform. According to the Fourier central slice theorem, [9], the 2D Fourier transform is equivalent to a Radon transform followed by a 1D Fourier transform along each projection. For the purposes of this method we will use the Radon transform as the forward operator $\mathcal{F}$, as there is no loss of quality from the 1D Fourier transforms. It has to be noted that the proposed method would work very similarly if the 2D Fourier transform was used for $\mathcal{F}$. An advantage of the Radon transform approach is that the grid of basis functions can be optimized to reconstruct only the basis functions which actually contain signal. The majority of medical images have zero intensities in the outer areas. This can be exploited by finding in Radon space the samples which have zero or very small values and then forming the corresponding line integrals to exclude grid points from the reconstruction.

Continuing from eq. 4 the data can be approximated by replacing f with its approximation $\tilde{f}$ (eq. 3):

$$\begin{aligned} g &= \mathcal{F}\tilde{f} \\ g &= \mathcal{F}Bp \end{aligned} \tag{5}$$

where B is the basis functions matrix and p is the vector of coefficients. Let $J = \mathcal{F}B$. The matrix J can be calculated analytically in the Fourier or the Radon space. The Fourier and Radon transforms of the Kaiser Bessel window functions can be found in [4].

The non-square matrix J is inverted using a generalized inverse.

$$p = J^{\dagger}g \tag{6}$$

A standard method for this inversion is the Moore-Penrose inverse $J^{\dagger} = (J^T J)^{-1}J^T$, which minimizes the least squares functional:

$$\Phi(p) = ||g - J(p)||_2^2 \tag{7}$$

To improve the stability of the problem we minimize the augmented functional $\tilde{\Phi}(p)$, as in [10], defined as:

$$\tilde{\Phi}(p) = \left|\left|\begin{pmatrix} g \\ 0 \end{pmatrix} - \begin{pmatrix} J(p) \\ \lambda I \end{pmatrix}\right|\right|_2^2 \tag{8}$$

where λ is a regularization parameter. This yields in the reconstruction formula:

$$p = (J^T J + \lambda I)^{-1} J^T g \tag{9}$$

The use of basis functions reduces the dimensionality of the problem. The number of unknowns in the linear system described above is decreased from the number of pixels to the number of coefficients of the basis functions. This produces a smoother reconstruction and therefore the introduction of an error between the approximated object $\tilde{f}$ and the ground truth object f. An analysis of this error can be found in [5] and [6].

3 Methods

3.1 Data acquisition

ECG gated data was acquired from a healthy volunteer. A total of 25 phases each with 208 radial profiles were collected using a five-element array receive coil. Each radial profile contained 352 samples. The data used to test the new reconstruction method was generated from this acquired data by undersampling to 8 radial profiles at each phase, resulting in a 26-fold acceleration compared to the original radial acquisition. A total of about 200 radial profiles can be collected within a single heart beat using a fast steady state free precession sequence. The sampling pattern was rotated by one projection for each phase, i.e. phase 1 data contained projections (1, 27, 53, ..., 183), phase 2 data (2, 28, 54, ..., 184), ... , phase 25 (25, 51, 77, ..., 207). This choice of undersampling spans the 180 degrees optimally both per phase and in total. The data used for the proposed method was 1D inverse Fourier transformed along each radial profile. This was further undersampled in the Radon space to 185 samples per projection.

All images were reconstructed to 128 by 128 pixel resolution. The grid resolution for the proposed method was 64 by 64 with the second degree Kaisser-Bessel basis functions (eq. 2) and the λ parameter was determined experimentally to be equal to 50. It has to be noted that the method is not sensitive to the regularization parameter.

3.2 Single coil

The first element of the coil array was chosen as an example since it contains most of the anatomy. The results are similar for other coils. The proposed reconstruction method was compared with a gridding reconstruction [11]. The root mean square (rms) errors were calculated by comparing the two methods with the gridding reconstruction obtained from the original fully sampled data. As we are not interested in the absolute intensity values but on their variation across the image, we compared the normalized images.

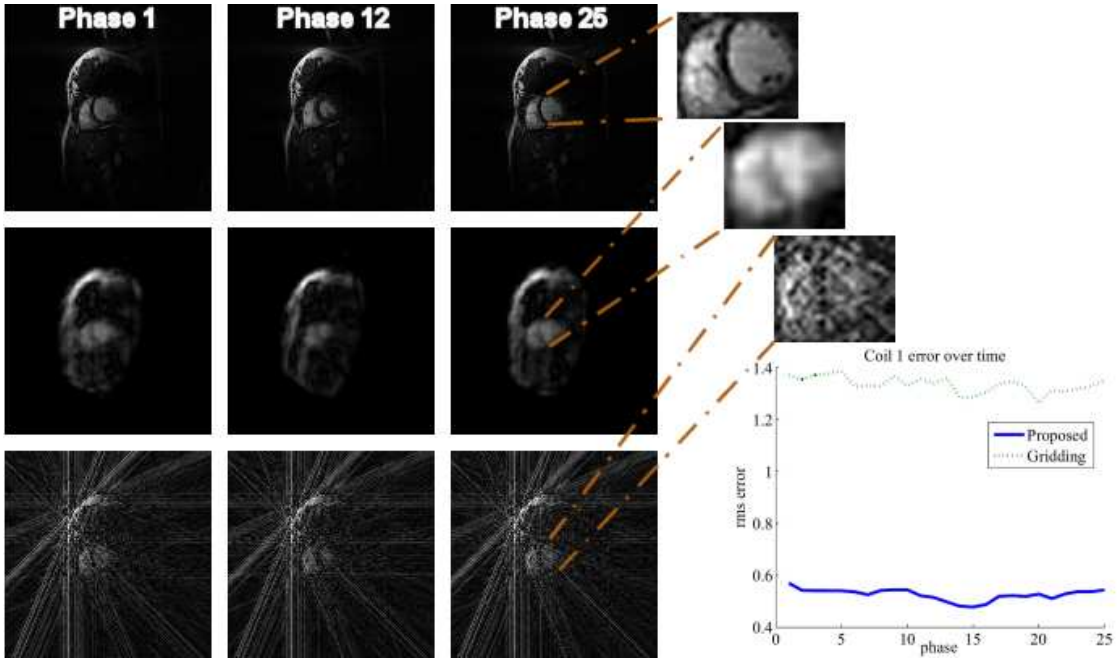


Figure 1. Single coil reconstructions with 8 radial profiles. Top row ground truth, middle row basis functions, bottom row gridding.

A similar strategy as described in section §3.1 was used for the various degrees of radial undersampling (4,13,16 profiles per phase).

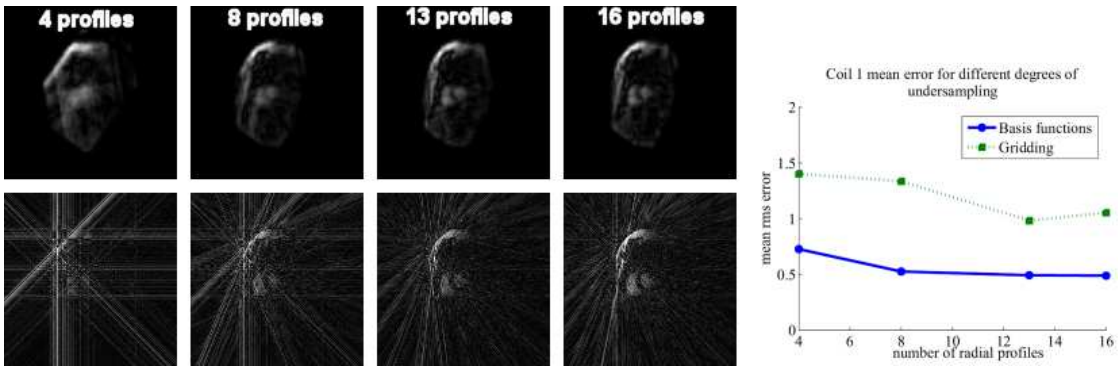


Figure 2. Single coil reconstructions of phase 12 with various degrees of undersampling. Top row basis functions, bottom row gridding.

3.3 Multiple coils

To reconstruct images from multiple receive coils using the proposed method, the operator $\mathcal{F}$ was changed to a Radon transform followed with a multiplication with the coil sensitivity matrix. As there was no body coil used for the collection of data, sensitivity matrices were calculated by dividing the time average image of each coil with the square root of the sum of squares of all the coil images, similarly to [12]. Each time averaged image was reconstructed using

all 200 radial profiles with a gridding reconstruction. For comparison the gridding reconstructed coil images were combined in a least-squares sense using the sensitivity matrices by solving the following linear system for each pixel:

$$\begin{aligned} C_{x,y} &= S_{x,y} f(x,y) \\ f(x,y) &= S_{x,y}^{\dagger} C_{x,y} \end{aligned} \tag{10}$$

where $S_{x,y}$ is a vector containing all the coil sensitivity values at pixel location (x, y), f is the image and $C_{x,y}$ is the vector containing the intensity value of each coil image at x, y. This method produces superior results when compared with the square root of the sums of squares of the coil images and it is computationally very efficient since it can be solved per pixel. Normalized images of the proposed method and the least squares gridding method were compared with the fully sampled images, which were reconstructed using the least squares gridding approach with the sensitivity matrices described previously.

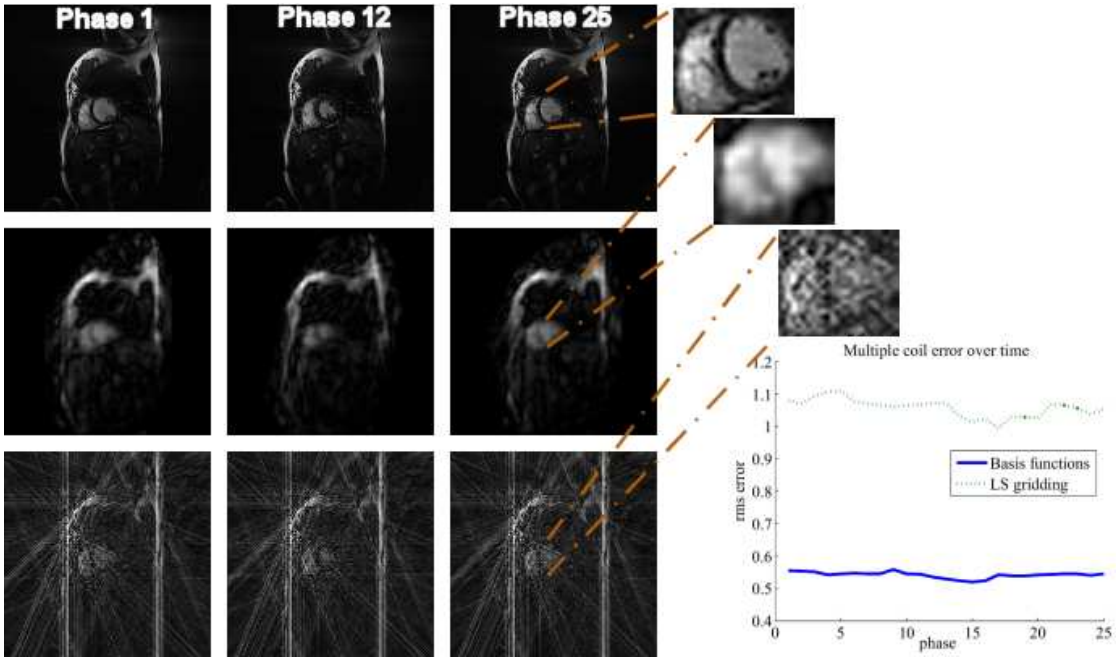


Figure 3. Single coil reconstructions with 8 radial profiles. Top row ground truth, middle row basis functions, bottom row gridding.

A similar strategy as described in section §3.1 was used for the various degrees of radial undersampling (4,13,16,26 profiles per phase). The coil sensitivity matrices used for the various undersampling patterns were the same as before.

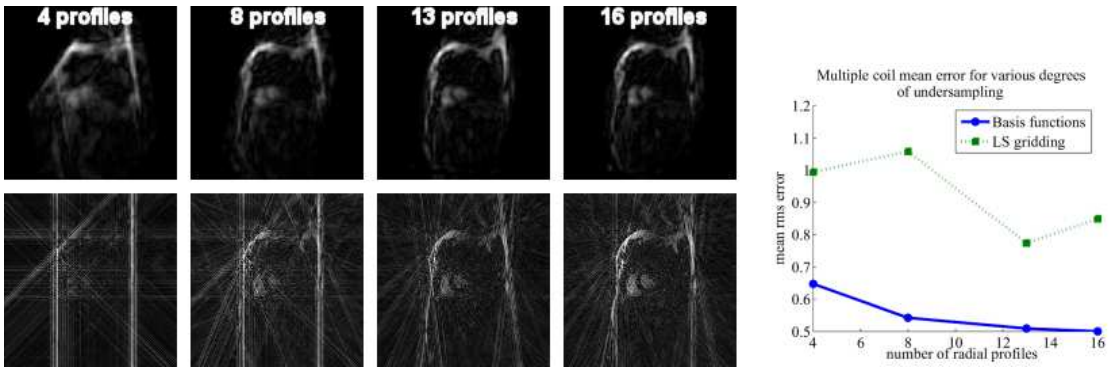


Figure 4. Multiple coil reconstructions of phase 12 with various degrees of undersampling. Top row basis functions, bottom row gridding.

4 Discussion

As demonstrated in the previous sections, the basis functions method performs better than gridding methods for various degrees of undersampling for both single and multiple coil data. Even though the gridding reconstructions can be marginally improved if data is smoothed, the superiority of the basis functions method still holds. This is though beyond the extend of this paper. The rms errors of the proposed reconstructions with 8 radial profiles ($\overline{rms}_s = 0.5268$, $\overline{rms}_m = 0.5427$) is actually lower than that of the gridding reconstructions with 16 radial profiles ($\overline{rms}_s = 1.0540$, $\overline{rms_m} = 0.8495$). To achieve the same quality of reconstructions with gridding methods a wider temporal window has to be used, which results in a lower temporal resolution. While increasing the temporal resolution, the left and right ventricles can still be clearly delineated in the proposed reconstructions. Further tests will assess the feasibility of the method in clinical environments. As the number of radial profiles per phase increases the gridding reconstruction improves, while the basis functions reconstruction converges to some error and the increase on quality is insignificant after that. The proposed method works with various degrees of undersampling and it could be used in adaptive data acquisitions, where the scanning pattern is adjusted according to the motion of the object of interest.

It also offers the ability to perform free-breathing data acquisitions as well as imaging patients with arrythmia. It can be used for imaging any dynamic object with high temporal resolution, where gated methods are simply infeasible. In these cases the coil sensitivity matrices can be calculated either from a body coil or from the basis functions images. Some of the temporal bandwidth could be sacrificed for collecting volume data. 3D reconstructions will be one of the subjects of future work. Comparing the basis functions method to SENSE [13] and k-t methods [14] will be also included in future work.

Acknowledgements

This work was partially funded by the EPSRC/MRC MIAS-IRC as part of the Intelligent Acquisition GC. I.K. would like to thank Michael Hansen for providing useful comments and the data used for these experiments.

References

1. V. Kolehmainen, S. Siltanen, S. Jarvenpaa et al. "Statistical inversion for medical x-ray tomography with few radiographs:ii. application to dental radiology." *Physics in Medicine and Biology* **48**, pp. 1465–1490, 2003.
2. R. Whitaker & V. Elangovan. "A direct Approach to Estimating Surfaces in Tomographic Data." *Medical Image Analysis* **6**, pp. 235–249, 2002.
3. D. F. Yu & J. Fessler. "Edge-perserving tomographic reconstruction with nonlocal regularization." *IEEE Transactions on Medical Imaging* **21**, pp. 159–173, 2002.
4. R. Lewitt. "Multidimensional digital image representations using generalized Kaiser-Bessel window functions." *Journal of Optical Society of America A* **7(10)**, pp. 1834–1846, 1990.
5. M. Schweiger & S. R. Arridge. "Image Reconstruction in Optical Tomography Using Local Basis Functions." *Journal of Electronic Imaging* **12(4)**, pp. 583–593, 2003.
6. J. J. Green. "Approximation with the radial basis functions of Lewitt." In J. Leversley, I. Anderson & J. C. Mason (editors), *Algorithms for Approximation IV*, pp. 212–219. University of Huddersfield, 2002.
7. H. Wendland. "Error Estimates for Interpolation by Compactly Supported Radial Basis Functions of Minimal Degree." *Journal of Approximation Theory* **93**, pp. 258–272, 1998.
8. M. Buhmann. "Radial Basis Functions." *Acta Numerica* pp. 1–38, 2000.
9. F. Natterer. *The Mathematics of Computerized Tomography*. Wiley, New York, 1986.
10. J. Varah. "A Practical Examination of Some Numerical Methods for Linear Discrete Ill-Posed Problems." *SIAM Review* **21(1)**, pp. 100–111, 1979.
11. J. Jackson, C. Meyer, D. Nishimura et al. "Selection of a Convolution Fucntion for Fourier Inversion Using Gridding." *IEEE Transactions on Medical Imaging* **10(3)**, pp. 473–478, 1991.
12. K. Pruessmann, M. Weiger, M. Scheidegger et al. "SENSE: Sensitivity Encoding for Fast MRI." *Magnetic Resonance in Medicine* **42**, pp. 952–962, 1999.
13. K. P. Pruessmann, M. Weiger, P. Bornert et al. "Advances in Sensitivity Encoding With Arbitrary k-Space Trajectories." *Magnetic Resonance in Medicine* **46**, pp. 638–651, 2001.
14. M. S. Hansen, C. Baltes, J. Tsao et al. "k-t BLAST Reconstruction From Non-Cartesian k-t Space Sampling." *Magnetic Resonance in Medicine* **55**, pp. 85–91, 2006.

Practical considerations for T_1-parametric mapping in colorectal and liver MR

Georgios Ketsetzis[a] and Michael Brady[b*]

[a] Wolfson Medical Vision Laboratory, Parks Road, Oxford OX1 3PJ

Abstract. T_1 mapping is a long standing problem in clinical MR, dating back to the early seventies. It was then first discovered that body tissues respond differently to magnetization and have different T_1 values. Although this fact created the possibility of tissue differentiation and tumour detection and characterization from MR images, such information is rarely utilized because of the practical difficulty in producing T_1 maps. Several different methods have been devised, all based on two fundamental T_1 mapping processes. We test the two fundamental processes for T_1 mapping of colorectal and liver MR mapping, concluding that these can fail depending on the quality of image data available. By modifying one of these two fundamental methods for T_1 mapping, we produce a new method for T_1 mapping that we apply to colorectum and liver MR. The new method proves more successful than conventional methods in providing reliable and higher quality T_1 maps that may be of real clinical use. The method is applied to produce the first ever good quality T_1 maps, known to us, of liver and colorectum MR.

1 Introduction

Two principal cancers that are responsible for millions of deaths every year are those of the liver and colorectum. Liver is one of the body organs most affected by metastatic cancer, with a total number of incidents of liver cancer being approximately 564.000 worldwide in the year 2000 [1]. Colorectal cancer affects around 10 million people each year worldwide, and colorectal cancer is now the second most common cause of cancer death in the western world. Clinical screening using MR is based on the use of high signal to noise ratio and high contrast to noise ratio images to enable diagnosis of disease, with or without the use of appropriate contrast agents when necessary. Therefore, the current practice of clinical MR assessment is based on signal intensity studies alone in an effort to determine the location, extent and stage of disease and to assess any course of treatment. Clearly, such an approach loses potentially valuable information that MR can additionally provide, specifically relevant to the nature of the underlying tissues, such as the T_1 and T_2 tissue relaxation times.

The T_1 relaxation time is known to increase in the presence of tumours and is reduced when the patient receives treatment [2]. The intravenous administration of gadolinium diethylenetriaminepentaacetic acid (Gd-DTPA) is also known to increase the signal intensity on T_1-weighted images by reducing the T_1-relaxation time. This feature is useful in distinguishing necrosis from active tumor and cystic lesions from solid lesions. T_1 and T_2 relaxation times have shown the potential to differentiate adenocarcinomas from muscle tissue but also the state of the adenocarcinoma, by differentiating the viable from the necrotic phase [3]. T_1 relaxation times are known to characterize different tissue types as the recent study by [4] demonstrated, in particular, abdominal tissues (Figure 1). Relaxation times provide tissue specific information that has potential of aiding clinicians in assessing the existence and stage of disease, but such information is rarely utilized in real clinical study. The reason for the rare use of T_1

a.

TABLE 1
Average T1 Relaxation Times at 1.5 and 3.0 T

Tissue	1.5 T: T1 Relaxation Time (msec)*	1.5 T: R^2 Value (%)	3.0 T: T1 Relaxation Time (msec)*	3.0 T: R^2 Value (%)	Difference (%)
Kidney					
Cortex	966 ± 58	0.999	1,142 ± 154	0.990	18†
Medulla	1,412 ± 58	0.997	1,545 ± 142	0.999	9
Liver	586 ± 39	0.995	809 ± 71	0.987	38†
Spleen	1,057 ± 42	0.998	1,328 ± 31	0.998	26†
Pancreas	584 ± 14	0.982	725 ± 71	0.976	24
Paravertebral muscle	856 ± 61	0.988	898 ± 33	0.988	5
Bone marrow (L4 vertebra)	549 ± 52	0.991	586 ± 73	0.994	7
Subcutaneous fat	343 ± 37	0.997	382 ± 13	0.999	11
Uterus					
Myometrium	1,309 ± 35	0.998	1,514 ± 156	0.999	16
Endometrium	1,274 ± 64	0.997	1,453 ± 123	0.998	14
Cervix	1,135 ± 154	0.998	1,616 ± 61	0.998	42
Prostate	1,317 ± 85	0.999	1,597 ± 42	0.998	21

* Data are mean ± SD.
† Difference is significant ($P < .05$).

b.

TABLE 2
Average T2 Relaxation Times at 1.5 and 3.0 T

Tissue	1.5 T: T2 Relaxation Time (msec)*	1.5 T: R^2 Value (%)	3.0 T: T2 Relaxation Time (msec)*	3.0 T: R^2 Value (%)	Difference (%)
Kidney					
Cortex	87 ± 4	0.993	76 ± 7	0.993	−13†
Medulla	85 ± 11	0.992	81 ± 8	0.996	−5
Liver	46 ± 6	0.992	34 ± 4	0.984	−26†
Spleen	79 ± 15	0.998	61 ± 9	0.996	−23†
Pancreas	46 ± 6	0.989	43 ± 7	0.977	−7
Paravertebral muscle	27 ± 8	0.925	29 ± 4	0.867	7
Bone marrow (L4 vertebra)	49 ± 8	0.997	49 ± 4	0.994	1
Subcutaneous fat	58 ± 4	0.995	68 ± 4	0.999	17†
Uterus					
Myometrium	117 ± 14	0.995	79 ± 10	0.993	−33
Endometrium	101 ± 21	0.987	59 ± 1	0.999	−42
Cervix	58 ± 20	0.993	83 ± 7	0.992	43
Prostate	88 ± 0	0.997	74 ± 9	0.995	−16

* Data are mean ± SD.
† Difference is significant.

Figure 1. T_1 and T_2 relaxation times for different tissue types. (a) displays measured T_1 for different abdominal and pelvic tissues by [4], clearly demonstrating the possibility of tissue discrimination using T_1 information. (b) displays measured T_2 values for different tissue types. Again, these values demonstrate that it is possible to differentiate tissues using parametric information. Images are courtesy of [4].

maps in diagnostic MR is due to the difficulty in obtaining them. Signal noise, poor SNR, poor CNR and motion

*corresponding author's email: geok@robots.ox.ac.uk

artifacts are the main causes of production of inaccurate or poor quality T_1 maps along with the disadvantages associated with increased data storage and the lack of optimized protocols for T_1 mapping to date. When extracted using conventional methods, Such maps could be of little if any use to disease assessment. This could explain why such maps have not been used to the best of our knowledge in clinical diagnoses for liver and colorectal cancer to date and no such maps have yet been produced for these organs.

The purpose of this paper is to demonstrate that conventional methods for the production of T_1 maps for liver and colorectum MR may not be effective and to examine the possibility of obtaining a new method for producing T_1 maps that may be of more value to the pathologists during clinical MR assessment of liver and colorectum disease.

2 Conventional methods for T_1 mapping in liver and colorectal MR

The T_1 and T_2 values can be obtained directly from the equation that provides the MR signal using non-linear least squares fitting of the given MR data to the MR signal equation. A model of the signal generated at a voxel by a gradient echo MR pulse sequence (FSPGR) is given by [5]:

$$S = g\rho e^{-\frac{TE}{T_{20}^*}} \sin\alpha \frac{1 - e^{-\frac{TR}{T_{10}}}}{1 - \cos\alpha e^{-\frac{TR}{T_{10}}}}, \tag{1}$$

where S is the measured signal, g is the scanner gain, ρ is the proton density, TE the echo time, TR the repetition time, α the flip angle and T_{10} and T_{20}^* are the longitudinal and transverse relaxation times respectively. The subscript 0 indicates that these maps are produced prior to the injection of contrast agent. For convenience in notation, we introduce the parameter k defined as:

$$k = g\rho e^{-\frac{TE}{T_{20}^*}}. \tag{2}$$

The conventional method of obtaining T_1 maps from MR data, which was originally introduced by [6], is to fit at least 3 low flip angle MR signal acquisitions (not necessarily at different flip angles) to a linear version of equation 1, given as:

$$\overbrace{\frac{S}{\sin\alpha}}^{Y} = \overbrace{e^{-\frac{TR}{T_{10}}}}^{A} \overbrace{\frac{S}{\tan\alpha}}^{X} + \overbrace{k(1 - e^{-\frac{TR}{T_{10}}})}^{B}, \tag{3}$$

Although this approach would produce T_1 maps that are increasingly less sensitive to noise (as the number of signal acquisitions increases), these maps will be increasingly more sensitive to mis-alignment errors of the fitted signals and the requirement for more scans may prolong patient inconvenience. Using the method of [6] and the optimized protocols suggested by [7], we mapped T_1 for a liver and a colorectum MR dataset obtained using the following parameter values: For liver MR: FSPGR sequence using TR = 8.9msec, TE = 4.2msec, flip angles 10 and 14 degrees, slice thickness = 4mm, field strength 1.5T. For colorectum MR, the following imaging parameters are used: flip angles 3, 10 and 17 and the same remaining imaging parameters and sequence as for the liver. Prior to T_1 mapping, the obtained images were registered using the method of [8]. These figures are displayed in Figure 2 a, b and c.

As we note from the obtained results (Figure 4, figures a), the produced T_1 maps are of poor quality due to the high contribution of signal noise (as there are not significant mis-registration errors between the images of each of the two sets) and it follows that the primary conventional method for T_1 mapping can produce poor quality T_1 maps. It is important to note at this point that the demonstrated failure is more due to the quality of the obtained data rather than the method of [6] itself. However, because that data was readily available to us from the Churchill hospital Oxford and without prior knowledge of the pre-processing steps needed to improve data quality for the purposes of T_1 mapping, the method of [6] practically fails. An alternative approach is T_1 mapping by using only two flip angle acquisitions, producing T_1 maps that are less sensitive to mis-alignment errors of the two signals but potentially more sensitive to signal noise. This analytical method for optimizing two flip angles for T_{10} measurement of the breast was introduced in [9]. In brief, the method can be summarized as follows: For a fixed voxel, and for two different flip angle gradient echo pulse sequences (such as FSPGR sequences), the signals are given by:

$$S_1 = k\sin\alpha_1 \frac{1 - e^{-\frac{TR}{T_{10}}}}{1 - \cos\alpha_1 e^{-\frac{TR}{T_{10}}}} \qquad \text{and} \qquad S_2 = k\sin\alpha_2 \frac{1 - e^{-\frac{TR}{T_{10}}}}{1 - \cos\alpha_2 e^{-\frac{TR}{T_{10}}}} \tag{4}$$

respectively. Dividing the above formulae, we obtain an analytic expression for T_{10} as

$$T_{10} = \frac{TR}{\ln\left(\frac{S_2\sin\alpha_1\cos\alpha_2 - S_1\sin\alpha_2\cos\alpha_1}{S_2\sin\alpha_1 - S_1\sin\alpha_2}\right)}. \tag{5}$$

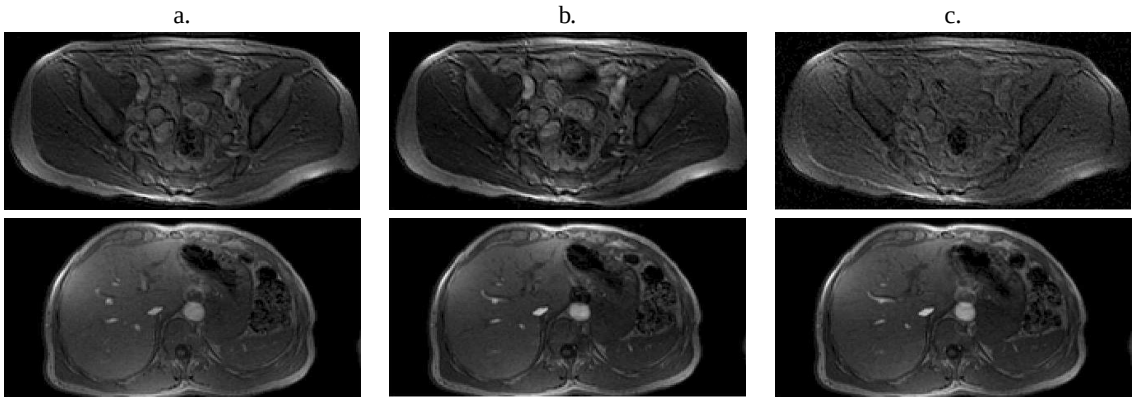


Figure 2. The original T_1 maps of colorectal (top row) and liver (bottom row) MR: Top row displays the original colorectal MR data corresponding to the flip angles 10 (a), 17 (b) and 3 (c) respectively which we use for T_1 mapping. Bottom row displays the original liver MR data corresponding to the flip angles 10 (a), 14 (b) and 14 (c) degrees. The later two acquisitions are made using the same parameters so as to improve the stability of the signal acquired at 14 degrees while mapping T_1.

Although this method avoids knowing the parameter k including its various parameters g, ρ and T_{20}^*, which would appear to be an advantage, the derivation of equation 5 assumes that the two noisy signals will each predict identical values for k as well as for T_{10}. This assumption fails in practice, as equation 5 turns out to be very unstable in the presence of signal noise (Figure 3). It appears therefore as if this method may not be appropriate for robust

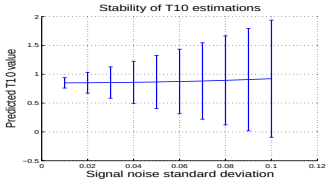


Figure 3. 95% confidence interval plot of the estimated value of T_{10} using the two flip angle formula, as a function of noise standard deviation. The flip angles used were $\alpha_1 = 3^o$ and $\alpha_2 = 10^o$ and the ground truth value of T_{10} is 0.85sec (average value of T_{10} for the breast in a magnetic field of 1.5Tesla). Notice that the computation of T_{10} using the 2-flip angle formula becomes very unstable with increasing noise. This simulation assumes Gaussian noise model, but because in real MR applications the noise model is Rayleigh distributed, the instability would be even worse.

estimation of T_1-maps of the liver and colorectum from MR images even when these have been registered.

We have produced T_1 maps by using both the aforementioned T_1 mapping methods and these appear in figures 4 a and b respectively. To produce figure 4 a we fit the three signals of figures 2 a,b and c for the colorectum and liver to equation 3 to map T_1 as it appears in figures 4 a. In figures 4 b we display the T_1 map of colorectum and liver MR by fitting the signals of figures 4 a and b only to equation 5. One of the reasons why the two methods have produced poor T_1 maps could have been the presence of the bias field. The bias field acts by multiplication on the image intensities, therefore at each image location the underlying image density S_i at location i is multiplied by a value b_i and the linearized form of the original signal in equation 3 takes the form:

$$\overbrace{\frac{S_i \star b_i}{\sin(\alpha)}}^{y} = \overbrace{e^{-\frac{TR}{T_{10}}}}^{a} \overbrace{\frac{S_i \star b_i}{\tan(\alpha)}}^{x} + \overbrace{g\rho e^{-\frac{TE}{T_{20}^*}}(1 - e^{-\frac{TR}{T_{10}}})}^{b}, \qquad i = 1, \ldots, n. \tag{6}$$

Because the bias field is practically identical for all slices in a data-set, this value of b_i is the same for all flip angles and is constant at each pixel (voxel) location i. The reason is that the bias field is caused because of inhomogeneities in the periodic RF pulse. But none of these is influenced by the flip angle or any other parameter used in the scanning process. The multiplicative action of the bias field has a geometric interpretation in the plane of the linearized signal equation: the bias field acts by multiplication equally on (x,y) in equation 6 which corresponds to a parallel translation of the ground truth linearized signal equation. As a result, the slope of the linearized signal equation, which characterizes uniquely the underlying T_{10} value, remains unchanged. Thus, the

T_1 values are not affected by the presence of the bias field. Similar arguments apply to the case of the two flip angle formula.

3 A new method for T_1 mapping: Application to liver and colorectal MR

Although the two flip angle formula may prove increasingly unstable in T_1 mapping, the fact that only two flip angle signal acquisitions are used and that the resulting T_1 maps are more stable to mis-alignment errors between the two signals than multiple flip angle acquisitions, suggests that there may be a possible modification of this formula that could increase the practical use of the resulting T_1 maps for clinical MR studies. In addressing this requirement, we introduce the following modification in a form of a theorem:

Theorem 3.1 *The value of T_{10} for GRE sequences at each voxel of interest is given up to a first order Taylor approximation as:*

$$T_{10} = \frac{TR}{\ln(\cos(\alpha_2)) - \frac{S_1}{S_2}\frac{\sin(\alpha_2)}{\sin(\alpha_1)}\left(1 - \frac{\cos(\alpha_1)}{\cos(\alpha_2)}\right)} \tag{7}$$

where S_1 and S_2 are the MR signals obtained at two flip angles α_1 and α_2.

Proof: The denominator in equation 5 can be written in the form:

$$\ln\left(S_2\sin(\alpha_1)\cos(\alpha_2) - S_1\sin(\alpha_2)\cos(\alpha_1)\right) - \ln\left(S_2\sin(\alpha_1) - S_1\sin(\alpha_2)\right) \tag{8}$$

which can be further written as:

$$\ln\left(\frac{S_2\sin(\alpha_1)\cos(\alpha_2)}{S_1\sin(\alpha_2)\cos(\alpha_1)} - 1\right) - \ln\left(\frac{S_2\sin(\alpha_1)}{S_1\sin(\alpha_2)} - 1\right) + \ln\left(S_1\sin(\alpha_2)\cos(\alpha_1)\right) - \ln\left(S_1\sin(\alpha_2)\right) \tag{9}$$

from where we observe that $\ln\left(S_2\sin(\alpha_1)\cos(\alpha_2)\right) - \ln(S_1\sin(\alpha_2)) = \ln\left(\cos(\alpha_1)\right)$. By introducing the approximation of the natural logarithm as: $\ln(1+x) = x - x^2/2 + x^3/3 + \cdots$ and taking the first order approximation, the first term of interest in the above denominator can be approximated by: $-(S_2\sin(\alpha_1)\cos(\alpha_2)/S_1\sin(\alpha_2)\cos(\alpha_1))$ and similarly, the second term in the expression 9 is approximated by: $S_2\sin(\alpha_1)/S_1\sin(\alpha_2)$. By putting all the above expressions together, we obtain the final formula for T_{10} which is exactly equation 7. ∎

In comparison to the T_1 maps obtained by using the original T_1 formula, the approximation introduced by equation 7, provides a substantial visual improvement of the previous results, where anatomical information is clearly present and provides more reliable T_1 maps. In spite of this advantage, the estimated values of the produced T_1 maps will only be an approximation of the true T_1 values which, in this particular case, are sufficiently good for our goal: to produce T_1 maps of colorectal and liver MR that will be of use to the pathologist for the diagnosis of disease and assessment of treatment.

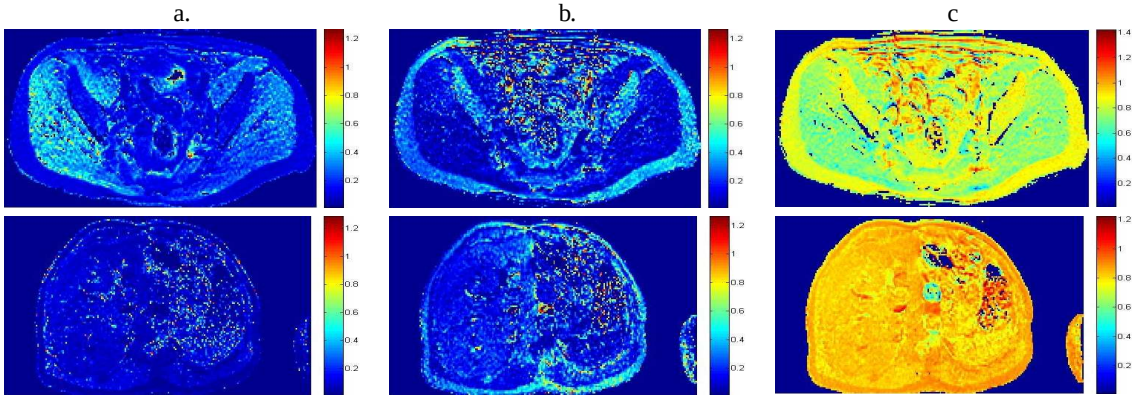


Figure 4. T_1 maps produced using the method of [6] (figures a), the method of [9] (figures b) and the proposed method (figures c). The noise levels of the original signal affect significantly the produced T_1 maps but the introduced approximation smoothes out the effects of noise, producing more plausible and more accurate T_1 maps: the T_1 values for colorectum and liver produced by the conventional methods are far below those known; the proposed method overcomes this.

The production of such T_1 maps for diagnosis of disease and treatment assessment has a potential advantage that is investigated in forthcoming work: The resulting T_1 maps can be segmented much more reliably using computer vision algorithms than the original intensity images because T_1 maps are not corrupted by the bias field; independently of this fact, the new method for T_1 mapping provides for the first time T_1 maps of colorectal and liver MR with a clear diagnostic potential.

4 Discussion and Conclusions

T_1 maps become increasingly important in image processing because of the important tissue specific information that these contain, which is not normally used in clinical MR. They have been shown to characterize different tissue types but they are rarely utilized because of the difficulty in producing them.

In this paper, we propose and test a method for overcoming the difficulty in producing quality T_1 maps of the colorectum and liver using conventional methods, by appropriately modifying a 2 flip angle T_1 mapping method. The modification appears to significantly improve T_1 mapping visually and qualitatively where conventional methods fail. We apply the results of the proposed method to obtain T_1 maps of colorectal and liver MR, producing such maps for the first time. The quality of the obtained T_1 maps raises the possibility of using them in aid clinicians assessments. In current work [10], we examine the segmentation of T_1 maps of colorectum and liver MR and provide quantitative comparison results against manual segmentation of T_1 and intensity data by pathologists and we also apply the proposed method [11] to the T_1 mapping of breast MR.

Acknowledgements

We thank Olivier Noterdaeme for providing the data and Anthony McIntyre for conducting the scans. We gratefully acknowledge General Electric for funding this project.

References

1. F. X. Boscht, J. Ribes, M. Diaz et al. "Primary liver cancer: Worldwide incidence and trends." *Gastroenterology* **127**, pp. 5–16, 2004.

2. T. Andersson, B. Eriksson, A. Hemmingsson et al. "Effect of interferon on T_1 relaxation times of liver metastases from endocrine gastrointestinal tumours." *Acta Radiologica* **29(1)**, pp. 21–25, 1992.

3. Z. Kovalikova, M. Hoehn-Berlage, K. Gersonde et al. "Age-dependent variation of T_1 and T_2 relaxation times of adenocarcinoma in mice." *Radiology* **164**, pp. 543–548, 1987.

4. C. M. J. de Bazelaire, G. D. Duhamel, N. M. Rofsky et al. "MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3.0 T:preliminary results." *Radiology* **230**, pp. 652–659, 2004.

5. R. Hashemi & W. G. Bradley(Jr). *Magnetic resonance imaging: The Basics*. Lippincott Williams & Wilkins, 1997.

6. R. K. Gupta. "A new look at the method of variable nutation angle for the measurement of spin-lattice relaxation time using Fourier transform NMR." *Journal of Magnetic Resonance* **25**, pp. 231–235, 1977.

7. P. Armitage, C. Behrenbruch, M. Brady et al. "Extracting and visualizing physiological parameters using dynamic contrast-enhanced magnetic resonance imaging of the breast." *Medical Image Analysis* **9(4)**, pp. 315–329, 2005.

8. D. Rueckert, L. I. Sonoda, C. Hayes et al. "Non-rigid registration using free-form deformations: Application to breast MR images." *IEEE Transactions on Medical Imaging* **18(8)**, pp. 712–721, 1999.

9. J. A. Brookes, T. W. Redpath, F. J. Gilbert et al. "Measurement of spin-lattice relaxation times with FLASH for dynamic magnetic resonance imaging of the breast." *British Journal of Radiology* **69**, pp. 206–214, 1996.

10. G. Ketsetzis & M. Brady. "Segmentation of parametric maps of colorectum and liver MR." *Submitted* 2006.

11. G. Ketsetzis & M. Brady. "A new method for T_1 mapping using FSPGR sequences: Application to breast MR." *Submitted* 2006.

Prostate Segmentation in MR Images Using 3D Shape Model Constrained Voxel Classification

P. D. Allen[a] and J. Graham[a] and D. C. Williamson[b] and C. E. Hutchinson[a*]

[a]ISBE, University of Manchester, Oxford Road, M13 9PT, UK,
[b]Department of Chemistry, University of York, YO10 5DD, UK

Abstract. Benign Prostatic Hyperplasia (BPH) is a non-cancerous expansion of the prostate, the progress of which can be quantified by measuring the relative volumes of the prostate's peripheral zone and central gland. Here we describe a method of automatic segmentation of both regions of the prostate from MR images using a combination of grey-level voxel classification and 3D statistical shape modelling.

1 Introduction

Benign Prostatic Hyperplasia (BPH) is a non-cancerous enlargement of the prostate which can cause constriction of the urethra and therefore obstruction of urinary flow. It affects 70% of men between the ages of 61 and 70, rising to 80% for men over 80 [1]. In 25% of men aged 80 symptoms are sufficiently severe to require surgical transurethral resection of the prostate (TURP), however this treatment has a high cost, morbidity (16%) and mortality (2.01%) [2] and so alternative treatments are sought.

Drugs such as finasteride can be used to treat BPH by shrinking the prostate, and evaluation of such candidate treatments requires a method of quantifying its effect. The current standard is Transrectal Ultrasound (TRUS) in which three orthogonal dimensions are measured and the volume is estimated using the formula for a prolate ellipsoid [2, 3]. Anatomically the prostate is divided into a number of zones: Peripheral (PZ), Central (CZ), Transitional (TZ), and fibromuscular. BPH primarily affects the TZ and so both the total Prostate (TP) and TZ volumes are measured using TRUS.

Tewari et al [2] have shown that the reduction in volume due to finasteride treatment over 12 months for the total TP and TZ are 8% and 27% respectively, however only the change in TZ volume correlates with improvement in urinary flow. For TRUS, intra-observer variability has been shown to be -18% to $+18\%$ for the TZ volume and -21% to $+30\%$ for the total volume [3].

Magnetic Resonance Imaging (MRI) is an attractive alternative to TRUS as it offers better definition of the prostate and is non invasive. MRI offers the possibility of accurately segmenting the prostate rather than assuming its volume from three orthogonal measurements. However, manual segmentation is time consuming, error prone and subjective, and so the goal of this project is to investigate the possibility of automatic segmentation of the appropriate regions of the prostate.

In MRI only two regions can be distinguished: the PZ, and what is referred to as the Central Gland (CG) comprising the remaining anatomical zones [4]. In cases of BPH the CG is mostly comprised of TZ due to the latter's expansion and so CG and TZ can be considered equivalent. Methods of segmentation [5] and registration [6] of just the outer prostate surface have been described for MR imaging, here we describe a method of both whole prostate and CG segmentation.

For this study we have used T2 weighted fat suppressed (T2FS) images as the CG/PZ contrast is enhanced in comparison with T2 or T1 weighting, and there is clearer separation of the prostate from surrounding tissue. The data were collected using a 1.5T Philips Gyroscan ACS MR scanner (software version NT5.3, Power Track 600, synergy body coil) from 22 patients with BPH. For each patient there are 50 axial slices with a thickness of 2mm and an in-plane resolution of 1.56mm.

Figure 1 shows a T2FS MR image of a prostate sliced in the axial, sagittal, and coronal planes. In T2 weighted images the PZ is generally brighter than the CG and in this case the two can be distinguished reasonably well.

Manual segmentation of the prostate is particularly difficult toward the superior portion where the seminal vesicles are very difficult to distinguish from the PZ, and toward the inferior portion where surrounding structures can become

*philip.allen@manchester.ac.uk

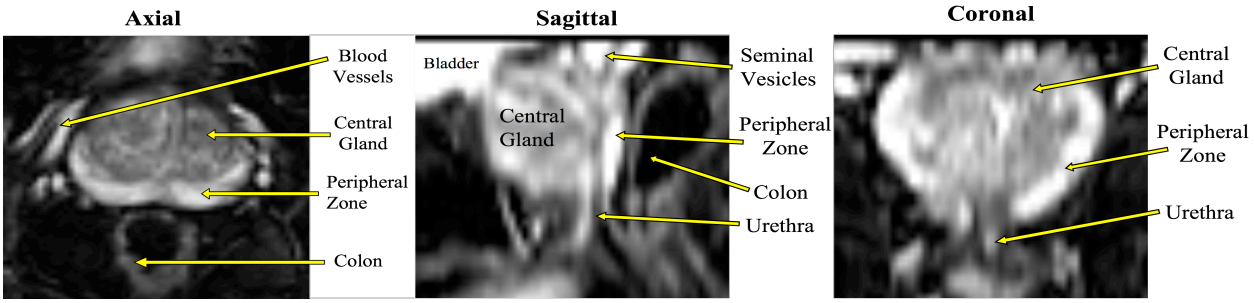


Figure 1. *The appearance of the prostate in T2 weighted Fat suppressed MRI sliced in three orthogonal planes.*

confused with the prostate and the prostate itself tends to bifurcate into two lobes. In the mid-section of the prostate blood vessels anterior to it can be confused with the PZ or CG depending on their relative intensity. The border between the PZ and CG can vary greatly from patient to patient depending on the severity of glandular enlargement. Figure 2 shows manual segmentation of the axial slice of the prostate, illustrating that the boundary between the regions is defined not only by the voxel values, but requires a model of expected shape.

2 Automatic Prostate Segmentation

In seeking a method of automatic segmentation, our approach is to formalise the two level process behind manual segmentation using grey-level voxel classification to make the initial coarse segmentation, and to fit a 3D point distribution model (PDM) [7] to this classified data to form the smooth spatial constraint.

2.1 Tissue Classification

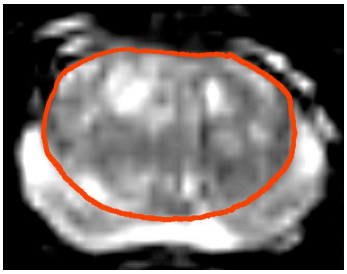


Figure 2. *Manual segmentation of an axial slice into PZ and CG.*

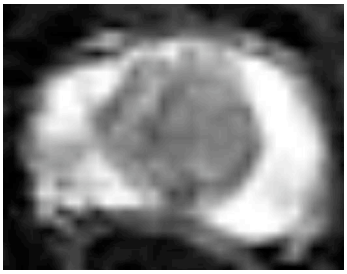


Figure 3. *Axial T2FS MR image of a prostate.*

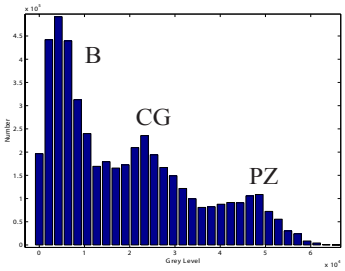


Figure 4. *The grey-level histogram of figure 3.*

Figure 5. *The result of applying grey-level tissue classification to figure 3.*

Figure 6. *The result of applying grey-level tissue classification to figure 2.*

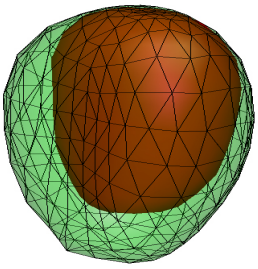


Figure 7. *The mean shape of the double surface 3D PDM of the prostate*

Figure 3 shows an axial slice of a T2FS images cropped close to the prostate. The histogram of grey-levels in this image is shown in figure 4. There are three distinct peaks in this histogram corresponding to PZ, CG and what we can consider background (B), suggesting that we can assume the image as being composed of three tissues. In an MR image a pure tissue type would produce a distinct grey level intensity with Gaussian distributed noise, so if the three tissue assumption holds we would expect the distribution of the histogram to be a sum of three Gaussians.

By fitting a three Gaussian-model to the histogram we can then calculate the class-conditional probability of each voxel

in the image belonging to each of the three tissue types [8, 9]. Figure 5 shows the results of using these probabilities to classify the image into PZ, CG and B. The example in figure 3 is particularly well suited to voxel classification. However in a less homogenous case such as that shown in figure 6 bright BPH nodules will be wrongly classified as PZ and the dark compressed regions of PZ wrongly classified as CG - thus a further spatial constraint is required.

2.2 3D Point Distribution Model

We are interested in fitting two surfaces: the Total Prostate (TP) and the Central Gland (CG) (see figure 7). The 22 images have been manually segmented to provide examples of each of these surfaces. To build a PDM from these surfaces requires a set of points on each surface which correspond across the data set and to achieve this we employ a method of automatic correspondence optimisation [10].

Using a leave-one-out evaluation of the ability of the PDM to represent the training data, it was found that using two separate PDMs for the TP and CG surfaces gave considerably better representation of the observed shape variability than using a single PDM of the two surfaces. This is because the 22 manually-segmented examples are not sufficient to adequately describe the variation in the spatial relationships between the two surfaces.

2.3 Model Search Method

2.3.1 TP Surface:

From the tissue classification (section 2.1) each voxel has three values associated with it - P_{PZ}, P_{CG}, P_B representing the probabilities that it belongs to PZ, CG, or B. For an example surface we can sum these probabilities for the voxels enclosed by the surface giving the quantities which we can call PZ_{in}, CG_{in}, and B_{in}. For fitting the whole prostate surface a sensible objective function would then be:

$$PZ_{in} + CG_{in} - B_{in} \tag{1}$$

We initialise the search for fitting all of the shape parameters by first fitting the pose of the average shape. Using the objective function in equation 1, and pose paramters only, the search space is fairly smooth. We are able to use simplex to find an initial configuration for shape search. Optimisation of the surface shape however presents a far more complex search space with many local minima, and so here a genetic algorithm is used.

2.3.2 CG Surface:

To fit the CG surface it is the PZ/CG border that must be emphasised in the objective function and this can be done in the following way: Create a candidate CG surface $C1$, then dilate that surface by one voxel to form a second surface $C2$, which, for a correct surface $C1$, should be outside the CG. If we sum the probabilities *on* the surfaces rather than *in* them we can form the values PZ_{onC2} and PZ_{onC1}. The difference between these values should be a maximum when the CG surface is on the PZ/CG border. As the PZ does not always extend round to the anterior of the prostate (see figure 1), we also need to find the CG/B border in this region. We therefore also calculate B_{onC2} and B_{onC1}. From a search point of view this objective function is spiky as candidate solutions near but not at the correct position are no better than surfaces further away, and so counting the voxel probabilities inside the surface is also necessary. Thus the CG objective function becomes:

$$(PZ_{onC2} - PZ_{onC1}) + (B_{onC2} - B_{onC1}) + CG_{in} - B_{in} \tag{2}$$

In this case the sums of probabilities *in* the surface are normalised by surface volume to make them the same order of magnitude as the sums *on* the surface which are normalised by surface area. The term ' $- PZ'_{in}$ is left out of the objective function because in some cases the there is considerable miss-classification of CG voxels as PZ.

Convergence of the CG fitting was not always successful starting from the mean shape. This difficulty was overcome by starting the search from a shape specified by a straightforward user interaction: the user marks four points on the middle slice roughly equally spaced around the CG, and selects the slices corresponding to the inferior and superior limits of the CG. This information is used to initialise the mean shape by stretching it in the X,Y, and Z axes and

Table 1. Fit Results (see text).

Surface	**Point Diff (mm)** $\mu(\sigma)$	**Volume Diff (%)** $\mu(\sigma)$
TP	4.1 (1.1)	11.1 (9.5)
TP (mid)	2.8 (0.82)	6.5 (5.4)
CG	3.1 (2.5)	11.9 (8.9)
CG (mid)	2.0 (0.6)	6.8 (8.5)

adjusting position in a simplex optimisation until the surface is as close to the user defined points as possible. The resulting surface is then used as a start point for a full GA optimisation of pose and shape.

2.4 Results

Prior to tissue classification and PDM fitting the image data was cropped manually around the prostate, as the full axial slices encompass the entire pelvic area in which the prostate is only a small region. The model fitting was tested in a series of leave-one-out experiments in which a surface model was built from the set of examples excluding the current example. GA optimisation was performed using MATLAB's genetic algorithm toolbox.

The results of fitting the TP and CG surfaces to each of the 22 patients are shown in table 1 as a mean point distance and percentage volume error. The effects of the anatomical ambiguity in the superior and inferior portions of the prostate (see section 1 can be reduced by only considering the mid-third of the cropped volume during shape and pose optimisation and the results of this are also included in table 1. Naturally the TP and CG volumes for the mid-third of the prostate are meaningless in themselves, however the CG/TP ratio calculated for this region from manual segmentation has a strong correlation (r=0.97) with the CG/TP ratio for the whole gland (figure 8). This suggests that to measure the CG/TP ratio, segmentation of the more clearly defined mid-section of the prostate may be sufficient.

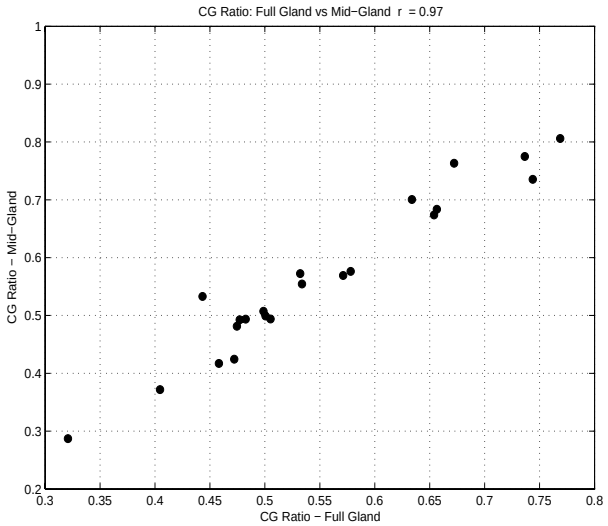


Figure 8. *The CG/TP ratio for the mid-gland plotted against the CG/TP ratio for the whole prostate.*

2.4.1 Repeatability:

Since the GA includes a stochastic element the same fit given the same data is not guaranteed. The magnitude of this variability can be estimated by repeating the fitting process 10 times and observing the variation in measured volume. On a subset of 10 of the patient group the results of this suggest standard deviations of 3% and 2% for the whole and mid-gland fits respectively.

3 Discussion and Conclusions

Table 1 demonstrates that in the majority of cases automatic segmentation results in Total Prostate and Central Gland surfaces that correspond accurately to the manual segmentation 'ground truth'. The key measure in this case is volume and even in cases where there is the greatest difference between automatic and manual segmentation these are comparable with the variation in volume estimates using TRUS. Automatic segmentation from MR images clearly has the potential to deliver precise estimates of volume change. Much of the discrepancy between manual and automatic segmentation arises in places where the boundary location is genuinely unclear, and the absolute nature of the ground truth is questionable. We intend to investigate the variability in manual measurement in these regions.

One of the more important measures to be derived is the ratio of volume of the Central Gland to Total Prostate. We have observed that this can be estimated reliably by only segmenting the more clearly defined central portion of the prostate. In a practical situation this may provide a solution to the more difficult cases.

References

1. A.M. Alam, K. Sugimura, H. Okizuka, J. Ishida, M. Igawa.: Comparison of MR Imaging and Urodynamic Findings in Benign Prostatic Hyperplasia. Radiation Medicine **18** (2000) 123–128
2. A. Tewari, K.Shinohara, P. Narayan: Transitional Zone Volume and Transitional Zone Ratio: Predictor of Uroflow Response to Finasteride Therapy in Benign Prostatic Hyperplasia Patients. Urology **45(2)** (1995) 258–265
3. A. R. Zlotta, B. Djavan, M. Damoun et al.: The Importance of Measuring the Prostatic Transition Zone: An Anatomical and Radiological Study. BJU International **84** (1999), 661–666
4. A. Maio, M. D. Rifkin.: Magnetic Resonance Imaging of Prostate Cancer: Update. Topics in Magnetic Resonance Imaging **7(1)** (1995) 54–68
5. Y. Zhu, S. Williams, R. Zwiggelaar.: Segmentation of Volumetric Prostate MRI Data Using Hybrid 2D+3D Shape Modelling. Medical Image Understanding and Analysis (2004) 61–64
6. B. W. Fei, A. Wheaton, Z. H. Lee et al.: Automatic MR Volume Registration and its Evaluation for the Pelvis and Prostate. Phy. Med. Biol **47(5)** (2002) 832–838
7. A. Hill, A. Thornham, C. J. Taylor.: Model-Based Interpretation of 3D Medical Images. British Machine Vision Conference (1993) 339–348
8. D. C. Williamson, N. A. Thacker, S. R. Williams, M. Pokric.: Partial volume tissue segmentation using grey-level gradient. Medical Image Understanding and Analysis (2002) 17–20
9. K. W. Fleischer, D. H. Laidlaw, A. H. Barr.: Partial-Volume Bayesian classification of material mixtures in MR volume data using voxel histograms. IEEE Transactions on Medical Imaging **17(1)** (1998), 74–86
10. R. H. Davies, C. J. Twining, P. D. Allen et al.: Shape Discrimination in the Hippocampus Using an MDL Model. Information Processing in Medical Imaging (2003), 38–50

A Charged Contour Model for Cardiac SPECT Segmentation

Ronghua Yang[a], Majid Mirmehdi[a], David Hall[b*]

[a]Department of Computer Science, University of Bristol, Bristol BS8 1UB, UK,
[b]Department of Medical Physics and Bioengineering, Bristol General Hospital, Guinea Street, Bristol, BS1 6SY, UK

Abstract. This paper presents a new deformable model for object detection, based on charged particle dynamics and geometric contour propagation, applied to the segmentation of the left ventricle of the heart in SPECT images. The model detects object boundaries with a charged active contour that propagates under the influence of Lorentz forces in an image-based electrostatic field. We find positions of vector field divergence for automatic initialisation of the contour. Experimental results on a dataset of 160 images are evaluated against hand-labelled groundtruth data.

1 Introduction

Myocardial perfusion scintigraphy (MPS) is carried out routinely in hospitals around the world, with the most common form of image acquisition being Single Photon Emission Computed Tomography (SPECT). Cardiac SPECT images provide information on myocardial viability, perfusion, and function, and this helps determine the presence, location, extent and severity of coronary heart disease. The functional parameters that assist the diagnosis include cardiac end-diastolic and end-systolic volumes and left ventricular ejection fraction which are calculated by detecting myocardial borders of the left ventricle (LV).

Active contour models have played a vital role in medical image analysis ever since the parametric snake model was introduced [5]. They have been used in cardiac SPECT segmentation with various modifications, e.g. see [2, 6]. However, they still suffer from problems such as edge leakage and high sensitivity to contour initialization.

In this work we propose a novel active contour model for LV segmentation in cardiac SPECT, namely the charged contour model (CCM). This is motivated by the recently introduced charged particle model (CPM) [4]. CCM defines a geometric active contour based on charged particle dynamics. The charged contour deforms under the influence of Lorentz forces in an image-based electrostatic field, and provides accurate boundary information. We also propose an automatic initialisation approach based on vector divergence estimation, which helps towards the automatic detection of both interior and exterior boundaries simultaneously.

In Section 2, we introduce the basis of our proposed model, charged particle dynamics, along with the CPM model [4] and its drawbacks. The structure of CCM is then outlined in Section 3, followed by the automatic initialisation procedure in Section 4. Experimental results are presented in Section 5, and the last section concludes the paper.

2 Charged Particle Dynamics and CPM

In charged particle dynamics, a set of positively charged, freely moving particles is placed in a field that has a distribution of negative fixed charges, forming an electrostatic field. Each charged particle is then attracted towards the fixed charges under the influence of a mesh-to-particle (Lorentz) force. At the same time, all positive free charges are repelled by one another by the repulsive particle-to-particle (Coulomb) force.

Consider an electrostatic field with M negative fixed charges and N positive free particles, with the k^{th} fixed charge at grid position $\mathbf{R}_k$ with charge $e_k < 0$, and the i^{th} free particle at position $\mathbf{r}_i$ with charge $q_i > 0$. Then, the Lorentz force $\mathbf{F}_l$ (in the absence of a magnetic field) and Coulomb force $\mathbf{F}_c$ acting on the i^{th} free particle are defined as [4]:

$$\mathbf{F}_l(\mathbf{r}_i) = q_i \sum_{k=1,\mathbf{R}_k \neq \mathbf{r}_i}^{M} \frac{e_k}{4\pi\epsilon_0} \frac{\mathbf{r}_i - \mathbf{R}_k}{|\mathbf{r}_i - \mathbf{R}_k|^3} \quad \text{and} \quad \mathbf{F}_c(\mathbf{r}_i) = q_i \sum_{j=1,j\neq i}^{N} \frac{q_j}{4\pi\epsilon_0} \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3}, \tag{1}$$

where ϵ_0 is the electric permittivity of free space. These equations state that the Lorentz force $\mathbf{F}_l$ acting on each free particle is the sum of the forces imposed onto it by all the fixed charges in the electrostatic field, and that the Coulomb force $\mathbf{F}_c$ is the sum of the forces imposed onto the particle by all the other free particles and as this changes dynamically with distance between particles, it needs to be updated over each particle movement. These two forces can

*Email: ronghua@cs.bris.ac.uk, majid@cs.bris.ac.uk, david.hall@ubht.nhs.uk

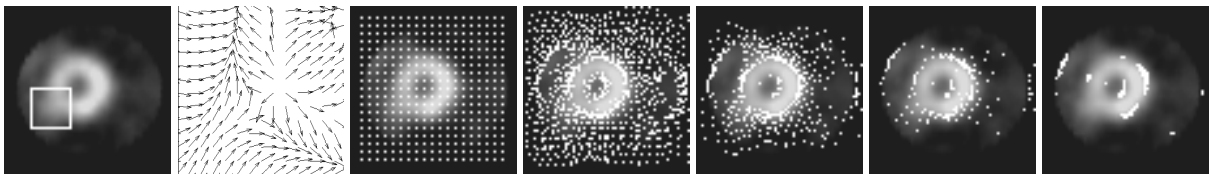


Figure 1. From left: original 64×64 cardiac SPECT image of a LV with a window highlighting a weak edge area, the normalized Lorentz force field within the windowed area, the initialized CPM [4] particles, and iterations of particle movements with the final result on the right.

play the roles of external and internal forces respectively in a deformable model for image segmentation. In CPM [4], the total force on each particle is a weighted sum of normalized Lorentz and Coulomb forces, reduced by a damping factor dominated by the particle speed at the previous step. As the weight of the Lorentz force is always larger than that of the Coulomb force, particles are primarily attracted towards object boundaries by Lorentz forces while Coulomb forces advance the particles along the boundaries. Thus, the boundaries can be recovered given proper dynamic particle addition and deletion during particle movements.

The CPM model [4] benefits from initialisation that is largely insensitive to placement. Nevertheless, it is very computationally intensive as (a) particles have to advance along the boundaries until they are evenly distributed (around the boundaries), and (b) particles are added and deleted dynamically at each iteration. Although a damping factor is used to reverse the direction of acceleration when a particle crosses an edge, the particle will still move as long as its speed is not exactly zero, and therefore oscillations occur at the boundaries and particle convergence needs to be detected by some criterion. Above all, CPM can not guarantee closed contours, which inevitably results in gaps in the recovered object boundaries particularly if the object is occluded or has weak edges, as shown in the example of a 64×64 cardiac SPECT image in Fig.1 (larger versions of all images in the paper can be viewed online[1]). The highly diffused edge area (see square on the original image) is significantly influenced by the stronger edges nearby, hence its Lorentz forces, especially those on the edge center, show tendency towards those stronger edges. As the Lorentz forces always dominate the direction of particle movements, particles which have arrived at the weak edges will continue moving to the stronger edges with the weak edge left unmarked. CPM therefore can not successfully deal with the SPECT images in our application as shown in Fig.1.

3 Proposed Charged Contour Model (CCM)

We propose a model built on the basis of charged particle dynamics combined with active contour evolution. This model detects objects starting with a positively charged active contour that propagates in an image-based electrostatic field, distributed with negative fixed charges, under the influence of Lorentz forces. The Coulomb force becomes redundant in our CCM model as repelling particle forces are no longer necessary. Instead, we use geometric curvature flow for model regulation, implemented using a level-set representation. CCM performs better than CPM in terms of (a) faster convergence, (b) steady equilibrium state, (c) less sensitivity in highly-textured images and (d) providing boundaries that are continuous closed contours.

A positively charged contour is placed in an image-based electrostatic field in which the fixed charge at every pixel position is computed as $e_k = -f(x_k, y_k) \leq 0$, for $k = 1, .., M$, where $f(x, y)$ is the image edge map and M is now the size of the image. Then, the Lorentz force field is obtained using the equation for $\mathbf{F}_l$ in (1). As the Lorentz force decays with squared distance, it produces weak flows in homogeneous areas and stronger flows closer to edges. This is an undesirable feature since, driven by these forces, the active contour can hardly move in homogeneous areas, whilst moving fast in edge areas where it is likely to oscillate around edges. As a result, we perform a weighted normalisation of the Lorentz force field to speed up the model convergence, such that the normalised field has strong flows in homogeneous areas and weak, or close to zero, flows near edges which finally stop the contour. At edge locations the strength of the flows are already drastically reduced due to the counteractions among the surrounding fixed charges. This is a good indication of where the real edges are located. Therefore, our edge-preserving weighted normalisation process is built on the information supplied by the edge map:

$$\mathbf{F}'_l(\mathbf{r}_k) = \frac{\mathbf{F}_l(\mathbf{r}_k)}{|\mathbf{F}_l(\mathbf{r}_k)|} exp(-|e_k|) \tag{2}$$

where e_k is the fixed charge located at $\mathbf{r}_k$ in the edge map. As $|e_k| \to 0$ in homogeneous areas, then $exp(-|e_k|) \to 1$, and the normalised Lorentz force $\mathbf{F}'_l(\mathbf{r}_k)$ has maximum magnitude tending to unity. When $|e_k|$ increases in edge

[1] *http://vision.cs.bris.ac.uk/AC/CCM/* (includes larger versions of all images, as well as extended comparative results)

regions, $exp(-|e_k|)$ scales $\mathbf{F}'_l(\mathbf{r}_k)$ inversely proportional to the edge strength $|e_k|$ and so $\mathbf{F}'_l(\mathbf{r}_k)$ reaches its minimum at the strongest edge location. Thus, the normalised Lorentz force field has the strongest vector flows in homogeneous regions which start to attenuate smoothly when entering edge neighbourhoods. Fig. 2 shows an examples of an original and normalised Lorentz force field of a simple shape with both interior and exterior boundaries along with a close-up look into both fields. As can be seen the normalised field $\mathbf{F}'_l$ is much more functional than the original $\mathbf{F}_l$ field in homogeneous or edge regions in achieving the objective of speeding or steadying the contour respectively.

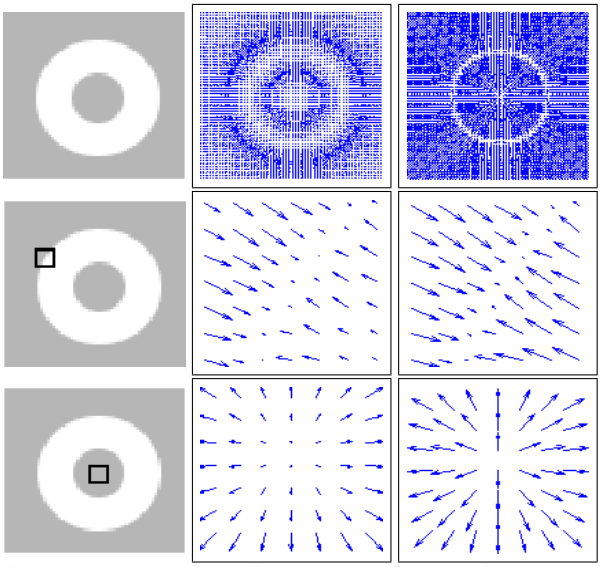


Figure 2. Top: Original image, its pre and post-normalised Lorentz fields, Middle and Bottom rows: close-ups on edge and homogeneous regions with pre and post-normalised fields respectively.

We then use curvature flow for contour regularisation based on [1] with our contour evolution function derived as:

$$\frac{\partial C}{\partial t} = \alpha g(x,y)\kappa\mathbf{N} + (1-\alpha)(\mathbf{F}'_l \cdot \mathbf{N})\mathbf{N} \tag{3}$$

where κ denotes the curvature flow, $g(x,y) = (1+f(x,y))^{-1}$ is a stopping term, $\mathbf{N}$ denotes the contour inward normal, and α is a positive constant that balances the contribution from curvature flow regularisation and the Lorentz force attraction. The first term regulates the contour. The orientation of the Lorentz force field at a point is towards strong nearby edges and thus every point of the charged contour will be attracted towards these edges and stop there, whilst being maintained as a closed contour. We embed the charged contour in a level-set so that it can propagate in the normalised Lorentz force field with topological flexibility:

$$\frac{\partial u}{\partial t} = \alpha g(x,y)\kappa|\nabla u| - (1-\alpha)\mathbf{F}'_l \cdot \nabla u \tag{4}$$

where u denotes the level-set. As the geometric contour always propagates in the direction of its normal $\mathbf{N}$, it is unable to move when the Lorentz forces are tangent to the contour. We compensate for this by adding an extra adaptive balloon force similar to Paragios et al. [7]. Hence the final formulation of CCM becomes:

$$\frac{\partial C}{\partial t} = \alpha g(x,y)\kappa\mathbf{N} + (1-\alpha)\{(1-\gamma(\mathbf{F}'_l \cdot \mathbf{N}))(\mathbf{F}'_l \cdot \mathbf{N}) + \gamma(\mathbf{F}'_l \cdot \mathbf{N})g(x,y)sign(\mathbf{F}'_l \cdot \mathbf{N})\}\mathbf{N} \tag{5}$$

where $\gamma = exp(-\lambda|.|)$ is a zero-mean Laplacian function which balances the contributions from the Lorentz and extra balloon forces. When the Lorentz forces and the contour normal are close to orthogonal, the extra balloon force contributes to shrink or expand the contour based on $sign(\mathbf{F}'_l \cdot \mathbf{N})$. Otherwise, the contour propagates mainly under the influence of Lorentz forces. The final level set representation is:

$$\frac{\partial u}{\partial t} = \alpha g(x,y)\kappa|\nabla u| - (1-\alpha)\{(1-\gamma(F_N))(\mathbf{F}'_l \cdot \nabla u) + \gamma(F_N)g(x,y)sign(F_N)|\nabla u|\} \tag{6}$$

where $F_N = -\mathbf{F}'_l \cdot \frac{\nabla u}{|\nabla u|}$ is the component of $\mathbf{F}'_l$ along the contour inward normal $-\frac{\nabla u}{|\nabla u|}$. Fig. 3 shows the propagation of CCM on a synthetic image. Note, CCM does not need to be initialised completely exterior or interior to the objects. However, in practice, background noise and features dictate a more elaborate initialisation scheme (see next section).

Computation of the Lorentz force field based on (1) is a simple, yet intensive, method which leads to $O(M^2)$ complexity. Therefore, we use the Particle-Particle Particle-Mesh method, originally proposed in [3], for fast and accurate evaluation of Lorentz forces, as in the CPM model [4]. The reader is referred to [3] for details.

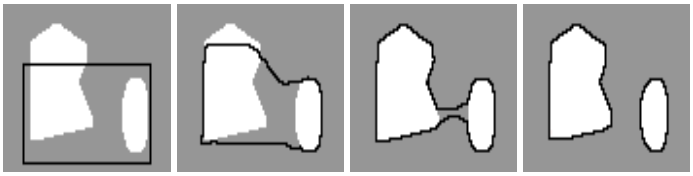


Figure 3. From left: the original image, initial CCM, and snapshots of contour propagation.

4 Automatic Initialisation

Here we propose a simple automatic initialisation method specifically designed for our cardiac SPECT application, although potentially it can be useful for other applications too. The divergence of a 2D vector field $\mathbf{F}$ is:

$$Div(\mathbf{F}) \equiv \nabla \cdot \mathbf{F} \equiv \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} \tag{7}$$

Therefore, the divergence values in a normalized image-based Lorentz force field, $Div(\mathbf{F}'_l)$, are usually negative in edge areas, and positive in homogeneous regions where the forces point at directions away from each other. We propose that if a point in the field where $Div(\mathbf{F}'_l(\mathbf{r}_k)) > 1$ is divergent enough to expand a contour, then it can be chosen as a candidate point for contour initialization. However, if we place an initial contour around every such candidate point, then contours propagating from both inside and outside an object will merge at the object boundaries due to the level-set formulation, and thus lead to false detection. Hence, we select the candidate points that are in the brightest regions of the image on the premise that these belong either to the object or the background. The contours then segment the object from *either the inside or the outside*. This works very well in practice for the cardiac SPECT images in which we used the top 20% of the brightest regions. This was determined empirically but is a very flexible threshold. Figure 4 shows an example of our automatic divergence point selection process, including a close-up of a divergence region and the entire divergence map.

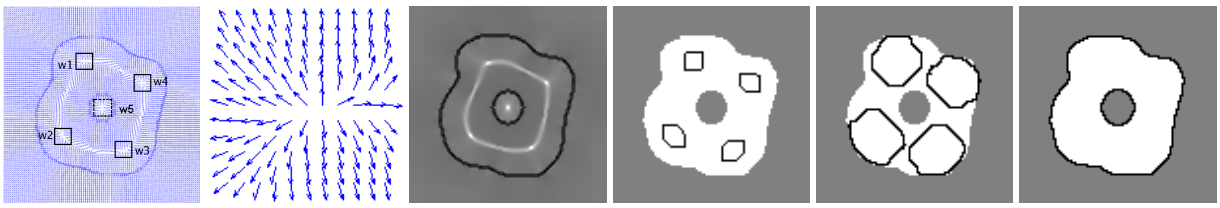


Figure 4. From left: Parts of the Lorentz force field where divergence is > 1 (i.e. w1-w5), close-up of w1, the divergence map with brightness proportional to divergence, automatic initialisation (w5 is discarded as it falls into the non-bright category), propagating CCM, final result.

5 Experiments

We applied our model to 64×64 pixel cardiac SPECT data of 20 individual patients, with 8 time frames per cardiac cycle. The parameters in the model were determined empirically and were kept constant throughout the experiments, i.e. $\lambda = 3$ and $\alpha = 0.1$. All patient studies were clinical studies acquired following standard protocols [8]. The results were evaluated against groundtruth data hand-labelled by an expert in SPECT imaging. Fig.5 shows snapshots of an evolving contour on an example LV image. Fig. 6 shows the initialisation and final contour results on a few more examples from our dataset, along with results from the CPM [4] model for comparison. The low resolution and largely diffused edges of the data impose a great burden on boundary detection, and CPM fails to recover most of the fuzzy edges (as described in Section 2) because particles inherently need to push along the boundaries. In comparison, CCM produces much better results as the contour propagates along its normal and can stabilise in the face of weaker edges.

We obtained 87.3% accuracy with $\sigma = 6.7\%$ in delineating the LV region across our dataset of 160 images based on a region overlap measure against the groundtruth data (itself subject to some human error). The best result was 98.8%, while the lowest was 60.7% where a large error resulted from extreme fuzziness of the edge areas due to the uptake of the radiotracer in the small bowel which can obscure the external border of the heart, particularly in the inferior wall.

6 Conclusions

We introduced a charged contour model for object detection based on charged particle dynamics and active contour propagation. An automatic initialisation method was also proposed to automate the model for convenient use in our

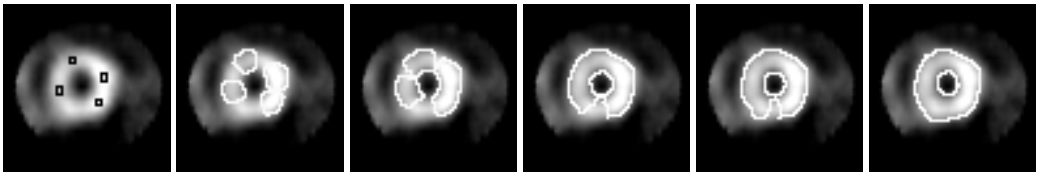


Figure 5. From left: Initial points of divergence, and snapshots of CCM detecting the internal and external boundaries of a LV on a short-axis cardiac SPECT.

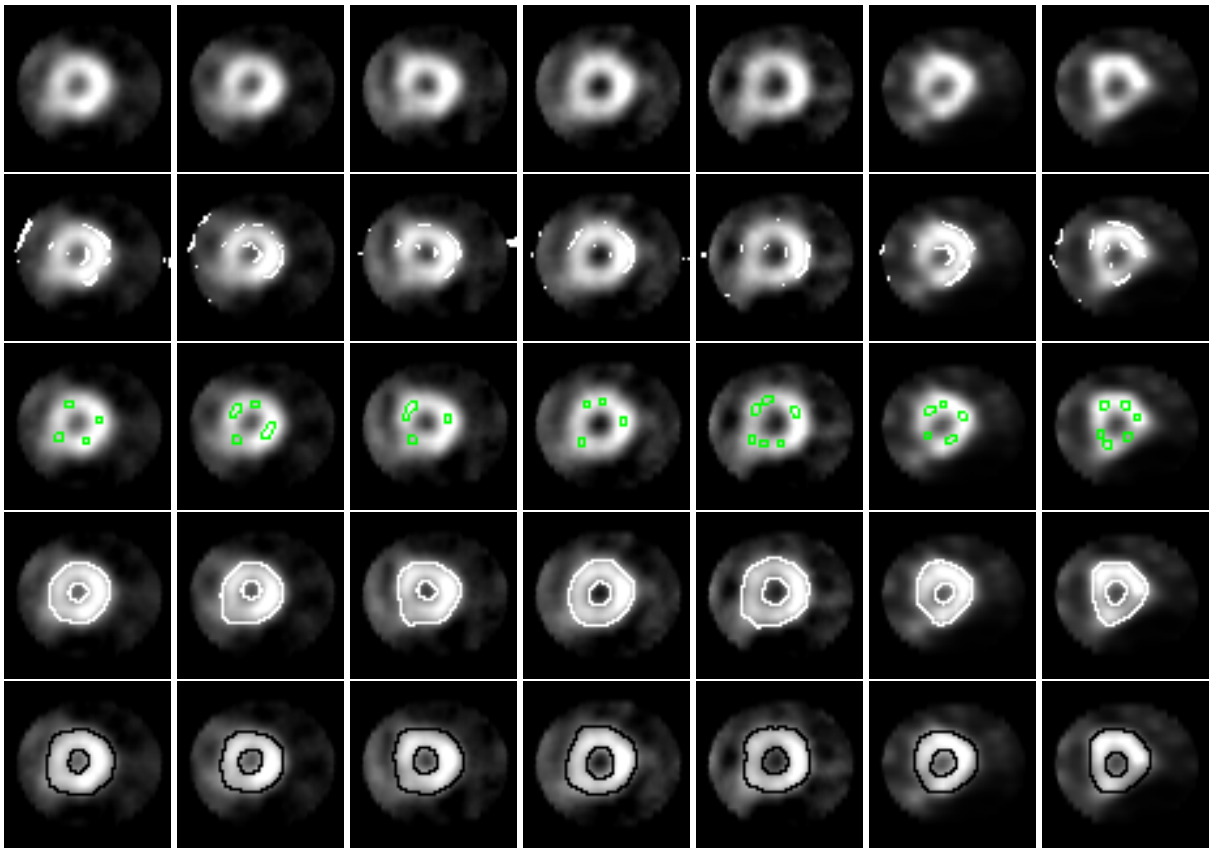


Figure 6. Mid-slice at 1st time frame from 7 different patients, from top to bottom: original image, CPM result, automatically initialized CCM, CCM result, and hand-labelled data.

cardiac SPECT application. A drawback of the current model is that it detects boundaries solely based on image gradients, and thus might lead to inaccurate definition of the internal and external borders of the LV, especially due to the fuzzy edges of SPECT data. Therefore, we aim to incorporate *a priori* knowledge constraints into the model such as constant myocardial volume for better LV segmentation and then motion estimation.

References

1. V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *ICCV*, pages 694–699, 1995.
2. E. Debreuve, M. Marlaud, G. Aubert, and J. Darcourt. Space time segmentation using level set active contours applied to myocardial gated SPECT. *IEEE Medical Imaging*, 20(7):643–659, 2001.
3. R.W. Hockney and J.W. Eastwood. *Computer Simulation Using Particles.* Taylor and Francis, 1988.
4. A.C. Jalba, M.H.F Wilkinson, and J.B.T.M Roerdink. CPM: A deformable model for shape recovery and segmentation based on charged particles. *IEEE PAMI*, 26(10):1320–1335, 2004.
5. P. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1:321–331, 1988.
6. R. Mullick and N. Ezquerra. Automatic segmentation of 3D cardiac SPECT imagery. In *IEEE SBEC*, pages 40–42, 1993.
7. N. Paragios, O. Mellina-Gottardo, and V. Ramesh. Gradient vector flow fast geodesic active contours. *IEEE PAMI*, 26(3):402–407, 2004.
8. T. Vakhtangandze, D.O. Hall, F.V. Zananiri, and M.R. Rees. The effect of Butterworth and Metz reconstruction filters on volume and ejection fraction calculation with $^{99}\mathrm{tc}^{m}$ gated myocardial SPECT. *BJR*, 78:733–736, 2005.

Simultaneous Registration and Landmark Detection

Sarah L Bond* and Michael Brady

Wolfson Medical Vision Laboratory, Dept. of Engineering Science, Oxford University, Oxford, OX1 3PJ, UK.

Abstract. We are developing a system for patient management in colorectal cancer, in which a difficult case of non-rigid registration, namely of pre- and post-therapy images, arises. Numerous non-rigid registration algorithms have been proposed in Medical Image Analysis, and we have applied several leading algorithms to our non-rigid registration problem; but with unpromising results. The fundamental reason appears to be that they lack with knowledge of the particular application. We propose a graphical representation of anatomical knowledge relevant for colorectal cancer, and of the ways in which this anatomy may be predicted to change as a result of chemo and radiotherapy. We show how we interleave this representation with an adaptive registration algorithm to make the non-rigid registration result both robust and accurate.

1 Introduction

A frequent patient management decision in colorectal cancer calls for the patient to undergo a course of neo-adjuvant chemotherapy, primarily to downstage a tumour, prior to surgical resection. This decision is usually based on small field of view T2-weighted MRI images [1], from which the tumour is first staged using the standard TNM scheme (T = primary tumour, N = lymph nodes, M = metastases) [2]. Following the course of chemotherapy, a second MRI is taken and assessed relative to that taken earlier, primarily to gauge changes in lymph node involvement, and tumour shrinking and invasion of the submucosa, surrounding muscle, muscularis propria subserosa or non-peritonealised perirectal tissues. This assessment requires that the MRI volume taken prior, be registered with that taken post, chemotherapy. The intrinsic complexity of colorectal anatomy, substantial variations to normal anatomy as a consequence of cancer, postural differences during imaging, and, most especially the substantial anatomical (and physiological) changes wrought by chemotherapy, mean that this turns out to be a very difficult registration problem, with substantial local deformations.

Due to the large scale changes involved in the deformations, many state of the art non-rigid registration algorithms still failed to give an acceptable registration result [3]. To improve the registration some prior knowledge needs to be incorporated, and this is often done using an initial segmentation. However using a segmentation on its own for the registration, although robust, is not inherently accurate. For an accurate registration, we would require a perfect segmentation, or else the errors in the segmentation can cause even greater errors in the registration. However it is impossible to achieve a perfect segmentation for the wide range of images that we are observing due to the vast differences in anatomy, size of patient, size of tumour etc.

It has recently been suggested that simultaneous segmentation and registration is a superior method of registering two images using the known segmented features within those images. Chen et al. [4] showed that by interleaving the segmentation and registration steps within a Bayesian framework, both the segmentation and registration results were improved. They were identifying tissue types in Contrast Enhanced Breast MRI images, whilst aligning the images that were taken over the course of 6 minutes. The images needed registering due to motion of the patient over this period of time. A maximum a-posteriori (MAP) estimation was used for the segmentation and registration stages, and a Markov Random Field was also incorporated as a noise reduction measure. The results showed a reduction in misclassification of pixels within the images.

The problem of aligning and segmenting contrast-enhanced breast images is however quite a different problem to that of aligning the Colorectal MR images, since the colorectal images are taken 6 months apart, and after extensive chemo and radiotherapy. As such the deformations are much greater and unpredictable, causing many generic, "off the shelf" registration algorithms to fail [3]. Although the problem of aligning the breast MR images is critical for diagnosis, the scale of the changes is much smaller, and the robustness problem is therefore less of an issue.

We have previously shown that incorporating shape knowledge can increase the robustness of the registration [3]. In this paper we show how we can incorporate the shape knowledge, or relational representation, into a simultaneous registration and landmark detection algorithm. Since we are not interested in segmenting every pixel in the image into a different class, but are only interested in key features in terms of the segmentation, we do

*Corresponding author: sarah@robots.ox.ac.uk

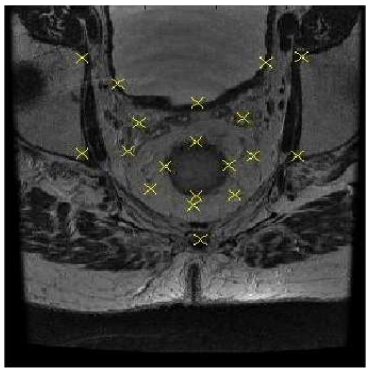


Figure 1. Initial landmarks found on the images.

not adopt a Bayesian MAP approach, but instead simply interleave landmark detection and registration within an adaptive registration framework.

2 Simultaneous Landmark Detection and Registration

2.1 Initialisation

An algorithm has already been implemented that finds landmark positions based on a relational representation of the anatomy visible within the images. These landmark points are features such as the colorectum, bone structures, mesorectum, and coccyx. They are found on the images using information about their appearance and positions. Full details of this method can be found in [5]. The landmarks themselves can be seen in Fig. 1. The method used gives a robust anatomical representation due to the relational position restrictions, but lacks accuracy due to the large variation in the images. Accuracy will be improved by interleaving the alignment of these features with the knowledge of the expected shape changes. Therefore, these landmark positions give the initialisation for this registration process.

2.2 Similarity Measure

The initial landmark estimates, though robust, are still liable to have an error of alignment, and are not guaranteed to be placed accurately. As such we need to update them. The crucial aspect is not so much their positions on the image, i.e. where exactly on the bladder the landmark is placed, it is more that the positions of the landmarks in the pre-treatment image correspond to the positions of the landmarks in the post-treatment image. We can look at the correspondence of the landmarks using Mutual Information, or Normalised Mutual Information. In this case, since the images are all T2 weighted MRI, we use Mutual Information, given by $C_{similarity}(A, B) = H(A) + H(B) - H(A, B)$, to provide a similarity measure between corresponding landmarks points. $H(A)$ and $H(B)$ are the entropies of image segments surrounding the landmark points on the two images and $H(A, B)$ is the joint entropy. We need to apply this similarity measure over a series of scales, for the areas surounding each landmark, to give a robust and accurate alignment. Hence the size of the image sub-block used to calculate the Mutual Information will be varied.

2.3 Registration

The method of Thin Plate Splines [6] is used to register the images. This provides a straightforward method by which to align corresponding landmark points. It is fast to calculate the warp field and so fast to calculate information about the warp field that can be used in the regularisation stage (Sect. 2.4).

The Thin Plate Splines are included in an adaptive registration framework similar to that developed by Park et al. [7]. This algorithm works iteratively over a series of scales, searching for the region of maximum misalignment. Park et al. define a mismatch measure based on Mutual Information $C_{similarity}$, given by $M = 1 - \frac{C_{similarity}(A,B)}{\min(H(A),H(B))}$. This identifies areas of high entropy and low local misinformation which makes for good placement of landmarks. The maximum mismatch can then be found and a landmark is placed and aligned. How-

ever, the initial problem we found with this algorithm was that these landmarks are often not within the region of interest. Since we have already defined our landmarks, and our goal instead is to align them, we use the minimum value of Mutual Information to decide which are our maximum misaligned landmarks, and we can consequently align these.

The alignment is based on a cost function that is due to the similarity between the landmarks, and the anatomy and physiology relationships on the two images. The cost function is minimised using simple gradient descent. We can use this simple method because local minima are unlikely within the small search region around each landmark point. We have already accounted for the large scale deformations by including these particular landmarks in the first place.

The registration operates across a series of scales from coarse to fine. As the scale becomes finer the sub-block size used to calculate the mutual information becomes smaller. The scale is refined when a threshold is met or when no further improvement can be made.

2.4 Regularisation

Shape knowledge was used in identifying the initial landmarks and this knowledge also needs to be incorporated throughout the registration, to ensure a fully simultaneous algorithm, making the most of all the information available.

There are two types of information that can be used. First, knowledge about anatomy, which involves information about the relative positions of the structures in the images. For example, the colorectum is always lower in the image than the bladder. The bone structures will be level assuming the images are rigidly aligned based on the hip positions, the coccyx is along the line perpendicular to that connecting the hips, from the midpoint of the hips and so on. These are included as series of weak constraints, that is, if they are **not** met then the cost function will increase substantially.

The other type of information that can be used concerns physiology. This is the knowledge of how the tumour and patient anatomy are likely to have changed from one scan to the next. For example, the tumour and its surrounding area may have contracted considerably, or alternatively there may be little change. Also, the bladder may be full in one image, and empty in another. We can represent both the anatomy and the physiology knowledge using a scalar representation of the warp field, the Jacobian. The Jacobian $J(A, B)$ gives an indication as to whether the tissue in expanding or contracting in a particular region [8]. A first estimate of this physiology change can be found by looking at the warp field of the initial landmarks and some calculated landmarks estimating the position of the mesorectal fascia and the colorectum, which will give an impression as to which regions are acting in which way.

A final cost measure of the anatomy and physiology is introduced based on the linear correlation between the expected and updated Jacobian representations: the two Jacobians are taken to be the two dimensions of a random variable $(J_{initial}(A, B), J_{update}(A, B)) = (X, Y)$, where $J_{update}(A, B)$ is the jacobian calculated from the current updated landmarks. The correlation coefficient, which is represented in the cost function is then $C_{physiology}(A, B) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$, where $\mu_X = E(X)$, $\mu_Y = E(Y)$, $\sigma_X^2 = E([X - \mu_X]^2)$ and $\sigma_Y^2 = E([Y - \mu_Y]^2)$ are the means and variances of the marginal distributions of X and Y, and σ_{XY} is the covariance of X and Y, given by $\sigma_{XY} = E([X - \mu_X][Y - \mu_Y]) = E(XY) - E(X)E(Y)$.

2.5 Final Algorithm

Putting all the landmark detection, adaptive registration and regularisation together the new algorithm can be illustrated as in Fig. 2. The algorithm looks very similar to the adaptive registration of Park et al. [7] except for the initialisation using the anatomy knowledge and the cost function which also incorporates that knowledge. The cost function to be minimised is

$$C = -C_{similarity} - \lambda C_{physiology}. \tag{1}$$

The choice of λ is important for the implementation of the algorithm as this indicates how much the prior information is used over the similarity and vice versa. For $\lambda = 0$ the anatomy and physiology knowledge is not

Do Find rigid features – hips and coccyx

Do Affine registration based on rigid features

Do Find initial estimates of anatomical relational landmarks, and initial Jacobian regularizer estimate

For I = large scale to small scale

Do

Identify Maximally misaligned landmark using MI

Optimize Landmark points to minimize cost function

While Cost function decrease > threshold

End For

Figure 2. Simultaneous Landmark Detection and Registration

Table 1. Comparison of the node misalignments (mm) for each of the 10 datasets, running the Park-Meyer algorithm on its own, the Park-Meyer algorithm with shape initialisation, and finally using the simultaneous algorithm.

Patient	Park-Meyer	With initialisation	Simultaneous
1	8.76	3.22	3.51
2	8.77	5.76	4.45
3	12.74	7.08	1.92
4	8.33	3.58	2.00
5	FAIL	7.01	3.48
6	8.76	3.88	3.48
7	5.06	4.06	3.65
8	FAIL	5.36	3.29
9	FAIL	6.36	5.00
10	7.59	3.77	3.48

incorporated at all, whereas for much larger values of λ the similarity measure has no effect. We ran the algorithm for large range of λ values on a series of 10 test images. The optimum value of lambda that we chose was $100\sqrt{10}$.

To improve the accuracy of the algorithm further in the mesorectal region of interest, some additional landmarks were added within the fat layer. Although they do not necessarily correspond to features or landmarks within the image, they will have a unique pattern due to any lymph nodes, blood vessels or artefacts such as streaking within that region. As such, they are useful to align to provide an even more accurate representation of the registration. The additional landmarks are added using the current points that are known such that we know that these landmarks will lie within the region of interest.

3 Results and Discussion

This algorithm gives a robust estimate of the registration for 10 corresponding pairs of unseen data slices (separate from those used in the training stage). That is, a clinically useful, sensible result is calculated for all cases. The precision of the algorithm was measured using corresponding lymph nodes, and the results can be seen in Table 1. Lymph nodes form distinctive landmarks that are unique to each patient. They are present in the mesorectum, which is the region of most clinical interest, and they are the features that clinicians look to compare in the pre and post therapy images. Hence it is crucial for the success of the registration that they are accurately and robustly aligned. However due to their uniqueness in individual patients it is not trivial to use them to drive the registration algorithm themselves. As noted in [3], prior to registration they can be typically misaligned by 21mm.

The algorithm was run using three sub-block sizes for 3 different scales, of 10, 5 and 2 pixels, and a step size of 1 pixel for the gradient descent. λ was chosen to be $100\sqrt{10}$. The overall mean node misalignment for all 10 data-sets is 3.42mm. Most enlarged lymph nodes are at least 5mm in diameter and hence an error of 3mm is an accurate registration result. All test were run in 2D, although the method is easily extendable to 3D.

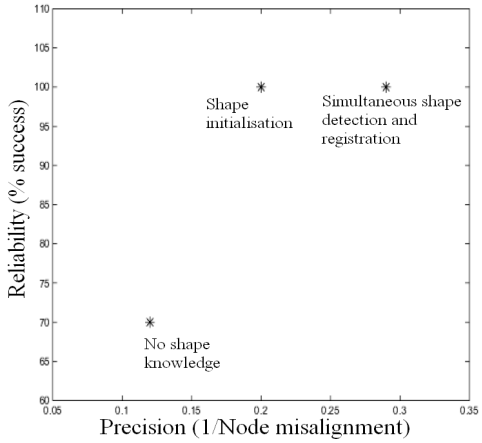


Figure 3. Reliability and Precision of non-rigid registration algorithms with and without the use of shape knowledge.

To compare, the adaptive registration algorithm with no shape knowledge input failed on 3 out of the 10 data-sets. That is corresponding lymph nodes could not be identified as they were misaligned by more than 14mm. This failure rate means that this would not be clinically useful. The precision of this algorithm was such that the mean distance between corresponding nodes was 8.6mm. Using the segmentation as an initial alignment, but not interleaving it simultaneously, the algorithm was much more robust and gave a clinically useful result in all cases. However the precision was such that the misalignment of the nodes was 5mm. These comparisons can be seen in Table 1 and Fig. 3. The points show the reliability, as percentage success over the 10 data-sets, against the precision. The precision is calculated as the average across the data-sets. The standard deviation for the results with just the Park Meyer algorithm is 3.65mm. With the shape initialisation only the standard deviation is 2.06mm, and with the simultaneous algorithm, it is 1.61mm.

Although the algorithm is more computationally costly when shape knowledge is incorporated, as there is the extra segmentation stage, the results show that when using this shape knowledge, the registration is 100% reliable on all the data-sets tested to date. Incorporating the shape knowledge into the simultaneous algorithm increases the computation further as the Jacobian is checked at each iteration. However, looking at the increased precision and robustness of the results by incorporating shape knowledge, it can be seen that this compensates for the increased computation.

4 Conclusions

We have shown that by introducing a simultaneous registration and landmark detection technique we can provide an extremely accurate and robust registration of colorectal cancer images pre and post treatment. These are data-sets which have large scale deformations, and cannot be registered using generic 'off the shelf' algorithms.

References

1. S. Bond & M. Brady. "Image analysis for patient management in colorectal cancer." *CARS* 2005.
2. G. Brown, C. Richards, J. Newcombe et al. "Rectal carcinoma: Thin-section mr imaging for staging in 28 patients." *Radiology* **211**, pp. 215–222, 1999.
3. S. Bond & M. Brady. "Non-rigid registration for colorectal cancer mr images." *CVBIA LNCS 3765* 2005.
4. X. Chen, M. Brady & D. Rueckert. "Simultaneous segmentation and registration for medical images." *MICCAI* 2004.
5. S. Bond. "Forthcoming thesis: Image analysis for patient management in colorectal cancer." 2006.
6. F. Bookstein. "Principal warps: Thin-plate splines and the decomposition of deformations." *IEEE PAMI* **11**, pp. 567–585, 1989.
7. H. Park, P. Bland, K. Brock et al. "Adaptive registration using local information measures." *MIA* **8**, pp. 465–473, 2004.
8. D. Rey, G. Subsol, H. Delingette et al. "Automatic detection and segmentation of evolving processes in 3d medical images: Application to multiple sclerosis." *IPMI LNCS 1613* pp. 154–167, 1999.

Combining Skin Pattern with Shape Analysis for Lesion Classification

Zhishun She [a, b], Y Liu [a] and A Damato [a]

[a] School of Science and Technology, NEWI, University of Wales, Wrexham, LL11 2AW, U.K.
[b] School of Electronic & Information Engineering, Southwest University, 400715, China.

Abstract: Skin line direction is extracted by skin pattern analysis as a measure of skin pattern disruption caused by the lesion. Shape analysis produces asymmetry, border irregularity and size of lesion. The four features are combined by the principal component analysis (PCA) for lesion classification. Test results show that classification on a set of 36 lesions containing 16 malignant melanomas has an area of 0.93 under the receiver operating characteristic (ROC) curve.

1 Introduction

As the survival rate of malignant melanoma (MM) depends on its thickness, diagnosis of MM at an early stage could reduce the risk of mortality and increase the chance of prognosis considerably. In order to achieve this, a computer automatic diagnosis (CAD) system used as a diagnostic aid in primary care is required.

In the CAD system it is known that the standard features for lesion classification are ABCD features, that is, asymmetry, border irregularity, colour variegation and size of lesion [1]. However the observation that skin patterning tends to be disrupted by malignant but not by benign skin lesions suggests that measurements of skin pattern disruption on simply-captured white light optical skin images could be a useful contribution to a diagnostic feature set [2]. Previous work using both skin line direction and intensity for lesion classification was encouraging [3]. However these new features have not been combined with the conventional features yet. All the features are required to be combined to enhance correct classification accuracy.

In the work described in this paper both skin pattern and shape analysis are conducted for skin lesion images. Skin line direction is extracted by skin pattern analysis. Three features are computed using shape analysis. They are asymmetry, border irregularity and diameter of lesion. Four features are combined with the principal component analysis (PCA) for lesion classification. The result of a classification test on a set of clinical skin lesions including 16 malignant and 20 benign lesions is encouraging.

2 Skin Pattern

Skin pattern can be produced by high-pass filtering [3]. As it is a weakly oriented texture, single directional analysis is useful in the effective identification and characterization [8]. From skin pattern image $G(m,n)$, the gradient vector $\nabla G(m,n) = [G_x(m,n), G_y(m,n)]^T$ is calculated where $G_x(m,n)$ and $G_y(m,n)$ are the gradients in horizontal and vertical directions. Then, an averaged tensor is computed on a local sub-image centred at pixel (i, j) with a size of *M*×*N* pixels. Skin line direction in this local area is determined by the eigenvector corresponding to minimum eigenvalue of this tensor. The angle of this eigenvector is determined by [3]

$$\phi(i, j) = \frac{1}{2}\tan^{-1}\left[\frac{2 f_{xy}(i, j)}{f_{xx}(i, j) - f_{yy}(i, j)}\right]. \quad (1)$$

Corresponding author: z.she@newi.ac.uk

where $f_{xx}(i,j) = \frac{1}{MN} \sum_{m=-M/2}^{M/2} \sum_{n=-N/2}^{N/2} G_x^2(i+m, j+n)$,

$f_{yy}(i,j) = \frac{1}{MN} \sum_{m=-M/2}^{M/2} \sum_{n=-N/2}^{N/2} G_y^2(i+m, j+n)$

and $f_{xy}(i,j) = \frac{1}{MN} \sum_{m=-M/2}^{M/2} \sum_{n=-N/2}^{N/2} G_x(i+m, j+n) G_y(i+m, j+n)$.

Next a snake-based edge detection technique is used to determine the lesion boundary [4]. The detected boundary segments the image into skin area A_s and lesion area A_l. Finally the average skin line directions in skin and lesion areas are calculated. Since the skin line directions are the directional data, the circular means rather than the arithmetic averages are utilized [5]. The circular mean of skin line direction in skin area is determined by

$$m_s = \begin{cases} \tan^{-1}(S_s / C_s) & \text{if} \quad S_s > 0 \quad \text{and} \quad C_s > 0 \\ \tan^{-1}(S_s / C_s) + \pi & \text{if} \quad C_s < 0 \\ \tan^{-1}(S_s / C_s) + 2\pi & \text{if} \quad S_s < 0 \quad \text{and} \quad C_s > 0 \end{cases} \tag{2}$$

where $C_s = \frac{1}{N_s} \sum_{(i,j) \in A_s} \cos\phi(i,j)$, $S_s = \frac{1}{N_s} \sum_{(i,j) \in A_s} \sin\phi(i,j)$ and N_s is the number of sub-images in the skin area. Similarly the circular average of skin line directions over lesion region m_l is computed. The absolute difference of skin line direction between skin and lesion areas is used for lesion classification.

As an example, the optical image of a malignant melanoma is shown in Figure 1 (a) and its skin pattern image is displayed in Figure 1 (b). Skin line direction was calculated in a local area with 16×16 pixels. The image of skin line direction is shown in Figure 1 (c). The circular means of skin line direction in skin and lesion areas were 49.25^o and 80.47^o, respectively. The absolute difference between them was 31.22^o indicating a significant disruption of skin line direction caused by the malignant melanoma. Classification using skin line direction was conducted [9] and the result is promising.

3 Shape Analysis

After the boundary is detected [4], a binary image, $A(k,l)$ is generated which satisfies $A(k,l) = 1$ if $(k,l) \in A_l$ and $A(k,l) = 0$ elsewhere. Shape analysis includes three steps to process $A(k,l)$. Firstly the central position of lesion is located. Then the orientation of lesion shape is determined. Finally the equivalent ellipse for the skin lesion is found.

Three features are calculated from shape analysis. They are asymmetry, border irregularity and diameter of lesion. Asymmetry is the behaviour of lesion shape about the major axis of the equivalent ellipse. One way to measure this is to fold the lesion outline about this axis, find the difference (the non-overlapping region) and calculate the percentage of this difference over the area of lesion, that is,

$$AS = \Delta T / T \times 100\% \tag{3}$$

where ΔT is the number of pixels in the difference area and T is the total number of pixels in the lesion region. Border irregularity is measured by the ratio of the square of the perimeter of lesion to the area of lesion. It is useful to lesion classification [10] and is computed by

$$B = \frac{P^2}{4\pi T} \tag{4}$$

where P is the perimeter of lesion boundary. The diameter of lesion is calculated by

$$D = 2a \quad (5)$$

where a is the semi-major axis of the equivalent ellipse. The scale is converted from pixels into millimetres (mm) by knowledge of image pixel parameters during image acquisition.

As an example, the optical image of a benign naevus (BN) is shown in Figure 2 (a) and its segmented binary image is displayed in Figure 2 (b) with its lesion centre (the crossing point of the major and minor axes), orientation (direction of the major axis) and the best-fit ellipse. The shape of lesion was rotated as shown in Figure 2 (c) so that the major axis was in the horizontal direction. Then top part was folded in the vertical direction. The difference between the folded part and the underneath part is shown in Figure 2 (d). Finally the three features were calculated using (3), (4) and (5), respectively. They were $AS = 16.78\%$, $B = 1.77$ and $D = 5.36$ mm.

4 Classification Results

The image set used in the experiment of this technique contains examples of several types of lesion including 16 melanomas and 20 compound or junctional naevi. The total number of cases to be acquired is limited by recruitment of patient volunteers in clinical practice. Histological diagnosis of skin lesions was conducted by dermatologist. In order to enhance the visibility of skin pattern side-lighting was used for image acquisition. The size of image is 230×350 pixels. It is chosen to have a reasonable skin area surrounding the lesions. Image pre-processing was applied to reduce artefacts such as hair and specular reflection.

Individual features which are described above were used for lesion classification. However the results are unable to achieve a high accuracy. A combination of features is needed for diagnostic decision. For lesion classification it is important to determine the sample size. It needs to satisfy $N > 3kp$ [6] where N is the sample size, k the number of classes and p the number of features. In this study we have two classes (BN and MM) and thus two features were extracted by PCA [7] in order to meet this requirement and then a linear discriminant analysis (LDA) was conducted for lesion classification.

The scatter-plot of the 36 skin lesions on the two-dimensional dominant feature space was computed and a receiver operating characteristic (ROC) curve was plotted using a linear classifier to segment the scatter-plot into classified-as-melanoma and classified-as-benign and varying the distance of the classifier from the origin. Figure 3 shows the ROC curve where the area under the curve is approximately 0.93, indicating an excellent classification result.

Lesion classification was also carried out using the leave-one-out (LOO) method to measure the classification performance. A linear classifier was trained using 35 samples and then a test was conducted using the excluded sample. The process was repeated 36 times, each time excluding a different sample. The classification result is shown in Table 1 which produces a sensitivity of 88%, a specificity of 95% and a correct classification rate of 92%.

5 Conclusions

Two complementary sets of features have been successfully combined to increase diagnostic accuracy. One feature set is extracted from the skin pattern which is characterised by the skin line direction. The other feature set is acquired by shape analysis of lesion which produces three features. They are asymmetry, border irregularity and diameter of lesion. Classification results indicate that combination of features from skin pattern and shape analysis enhances the classification results significantly, suggesting that combining skin pattern and shape analysis is promising in the CAD system. Future work is to investigate the gradient variability of skin pattern inside and outside the lesion.

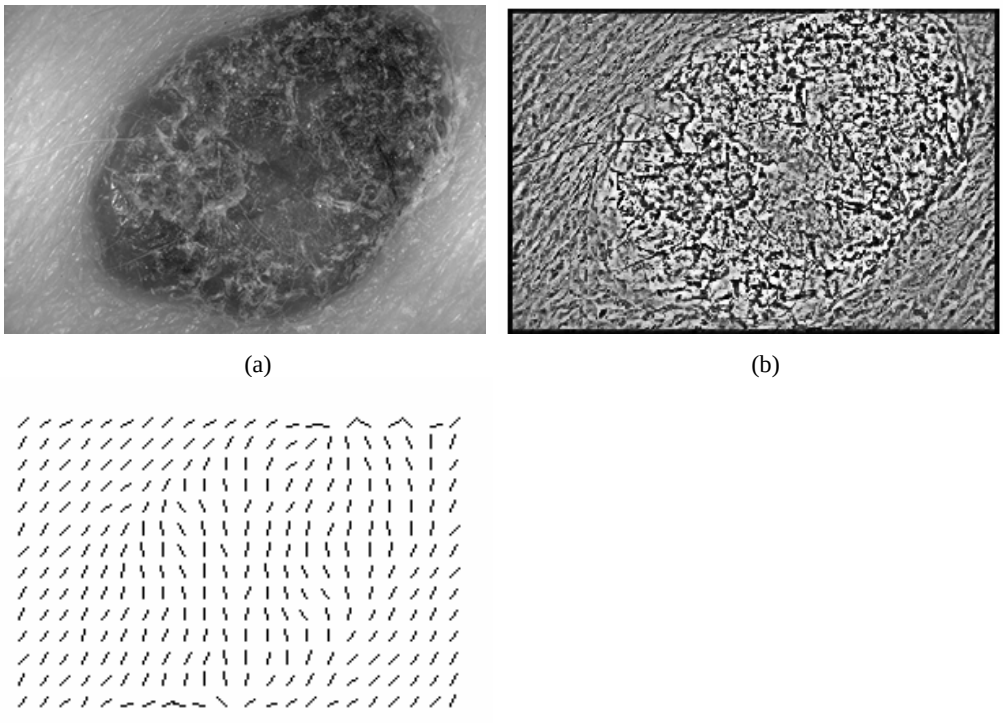
Acknowledgements

The optical skin lesion images were provided by M.Dickson, V.Wallace and Dr. J.Bamber of the Physics

Department, Clinical Research Centre, Royal Marsden Hospital, Sutton. Permission to use them is gratefully acknowledged. This work was funded by EPSRC grants GR/M72371 and GR/M72289.

References

1. R.M. Mackie, *An illustrated guide to the recognition of early malignant melanoma*, Blackwood Pillans & Wilson Ltd., Edinburgh and Department of Dermatology, University of Glasgow, 1986.
2. P.N.Hall, "Clinical diagnosis of melanoma", *Diagnosis and management of melanoma in clinical pratice*, pp. 35-52, Springer-Verlag, New York, 1992.
3. Z. She and P.J.Fish, "Analysis of skin line pattern for lesion classification", *Skin Research and Technology*, **9**, pp. 73-80, 2003.
4. Z. She & P.Fish, "Boundary detection of skin lesion using a fast snake algorithm", In *Proceedings of 16th biennial international EURASIP Conference*, pp. 295-297, 2002.
5. K.V. Mardia, *Statistics of directional data*, Academic Press, 1972.
6. D.H.Foley, "Consideration of sample and feature size", *IEEE Transactions on Information Theory*, **18**, pp. 618-626, 1972.
7. S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 2003.
8. A. R. Rao, *A taxonomy for texture description and identification*, Springer-Verlag, New York, Inc., 1990.
9. Z. She & P.Fish, "Skin lesion differentiation using skin line direction", In *Proceedings of Medical Image Understand and Analysis*, pp.169-172, 2002.
10. B.S.Airbisala and E.Claridge, "A border irregularity measure using hidden markov models as a malignant melanoma predictor", In *Proceedings of Medical Image Understand and Analysis*, pp.231-234, 2005.



(a) (b)

(c)

Figure 1 (a) original malignant melanoma image, (b) skin pattern and (c) skin line directions.

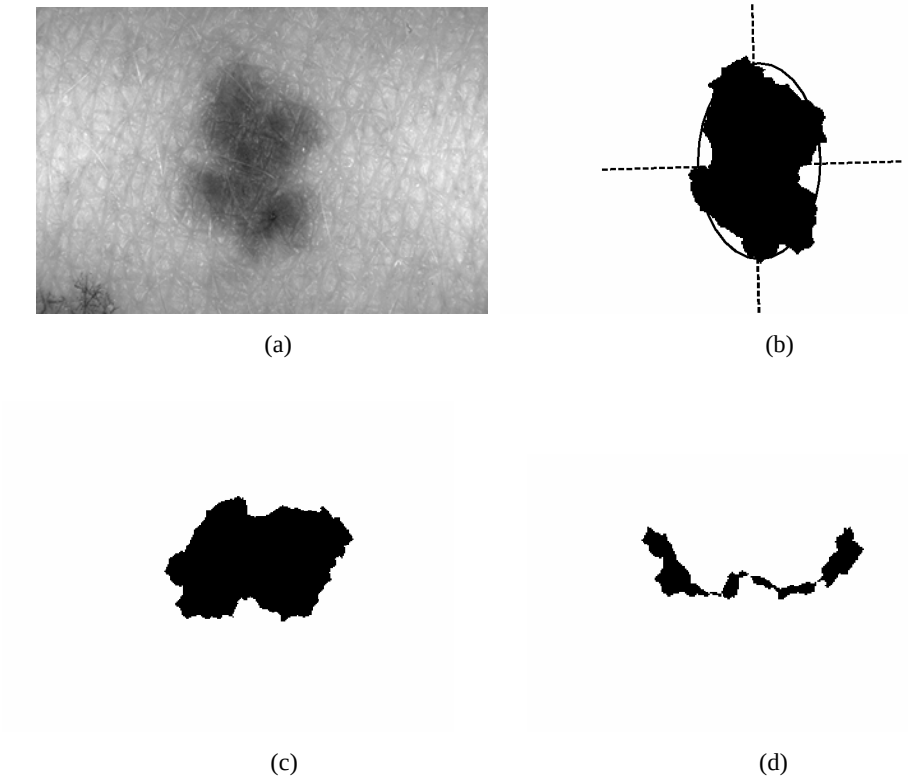


Figure 2 Shape analysis of benign naevus: (a) original image, (b) lesion shape with its centre, orientation and the best-fit ellipse, (c) rotated segmented image and (d) the difference.

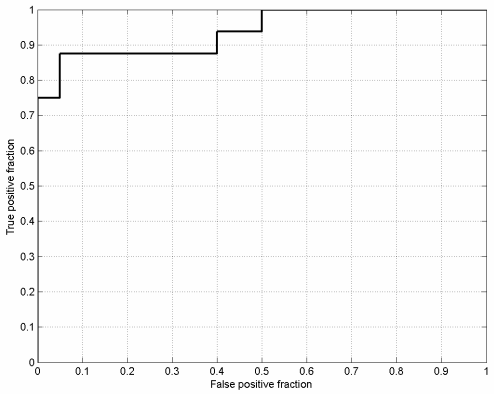


Figure 3 ROC curve for lesion classification.

Classified type

Actual type	MM	BN
MM	14	2
BN	1	19

Table 1 Classification result using LOO.

Retinal vessel detection using greyscale skeletons and adaptive contrast assessment

Alan D Fleming[a]*, Keith A Goatman[a], Peter F Sharp[a] and John A Olson[b]

[a]Biomedical Physics, Aberdeen University, Foresterhill, AB25 2ZD,
[b]Grampian Diabetes Retinal Screening, Woolmanhill Hospital, Aberdeen, AB25 1LD

Abstract. This paper shows how the detection of vessels in photographs of the human retina can be improved by adaptive contrast assessment. Automated analysis of retinal photographs has become an important issue due to the nationwide implementation of screening programmes for diabetic retinopathy, a disease that causes blindness if not detected early enough. Computer detection of the vessels in the images is one step towards a fully automated analysis of the image in terms of disease, image quality and comprehension of the anatomical structure. Although the larger vessels are distinct from their background, recognition of smaller vessels requires comparing their contrast with the contrast of the surrounding retina. We show that the false detection of image noise as vessels can be reduced by up to 60% by adaptive contrast assessment.

1 Introduction

Diabetic retinopathy screening programmes, based on digital photography, are being set up across the UK and aim to reduce the incidence of blindness in people with diabetes. All people with diabetes are invited for annual screening resulting in a large number of images to be analysed. It is widely agreed that computer analysis will aid the implementation of diabetic retinopathy screening programmes. Detection of retinal vessels plays an important role in this since (a) they are less likely to be confused with lesions of retinopathy if they have been separately identified, (b) the advanced forms of diabetic retinopathy affect the vasculature, (c) the visibility of small vessels is important for checking that image quality is adequate [1], (d) reliable detection of the optic disc and fovea requires initial detection of vessels and (e) image registration may be required [2]. Many previous studies have reported methods for retinal vessel detection, important examples of which are [3] and [4]. Good results for vessel detection with adaptive contrast enhancement are described in [6].

This paper shows that adaptive contrast assessment, similar to that described in [5] for microaneurysms, improves vessel detection. Retinal vessels were detected by generating a grey-scale skeleton. This is divided into segments which are classified as vessel or non-vessel. We tested the classification with and without adaptive contrast assessment.

2 Methods

2.1 Image pre-processing

The methods are described for images of the retina with a 45° field of view and of size approximately 1700×1600 pixels. The vasculature cannot be reliably identified in the red and blue planes of retinal photographs. Only the green plane shows the vessels with high contrast and this was extracted for vessel detection. The background intensity, denoted B, was estimated by applying a median filter to I with sufficient size (67×67) to be little affected by vessels up to 30 pixels calibre. A shade-corrected image, was generated with $I' = I/B - 1$ and was normalised for global image contrast by dividing by its standard deviation (sd), Figure 1(i),

$$I_S = I'/\mathrm{sd}(I'). \tag{1}$$

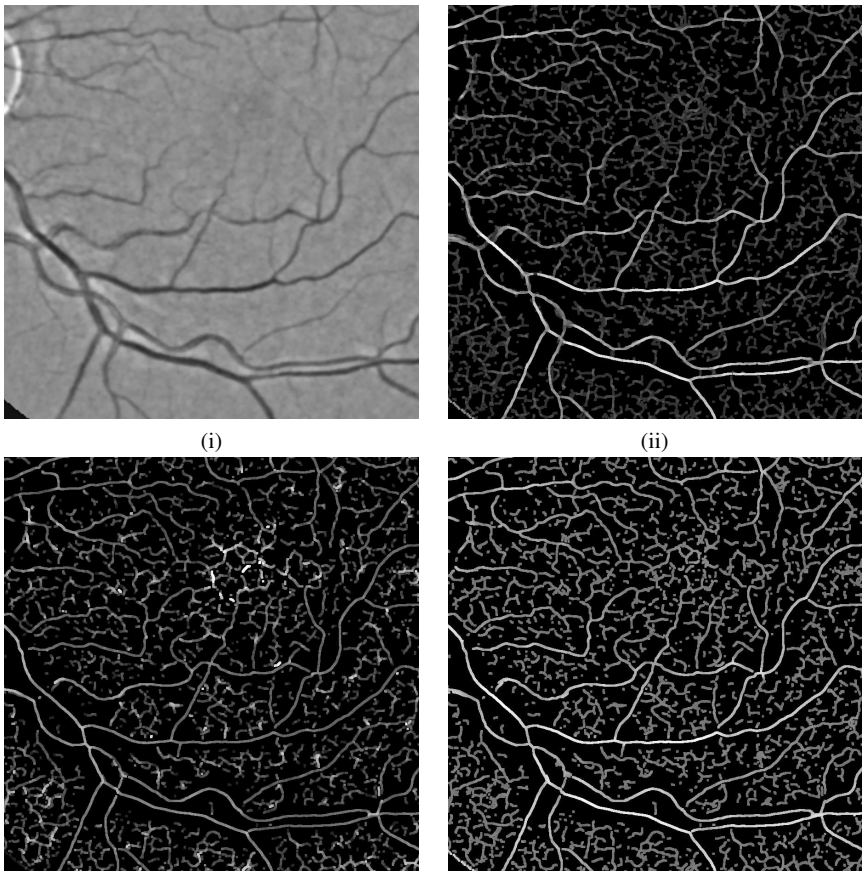
The value of I_S at a vessel is typically -6 to -1 while retina is typically -0.5 to $+0.5$.

2.2 The greyscale skeleton

A variation of the "min-max medial axis transform" of [7] was used. A sequence of dilations $J_r, r = 1, 2, \ldots 15$, were evaluated (with '$\oplus$' representing dilation) using

$$J_r = \begin{cases} J_{r-1} \oplus B_1 & \text{if } r \text{ is odd} \\ J_{r-1} \oplus B_2 & \text{if } r \text{ is even} \end{cases} \tag{2}$$

*Correspondence to A. Fleming, a.fleming@biomed.abdn.ac.uk



(i) (ii)

(iii) (iv)

Figure 1. (i) Extract from shade corrected image. (ii, iii & iv) Greyscale skeleton M, R and L.

where $J_0 = I_S$, $B_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $B_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. Repeated dilation alternately by B_1 and B_2 is equivalent to dilation by an octagonal structuring element, a much closer approximation to a circular disc than repeated dilation either by B_1 or by B_2. The greyscale skeleton was then evaluated (with '$\bullet$' representing closing) as

$$S_r = \begin{cases} (J_r \bullet B_1) - J_r & \text{if } r \text{ is odd} \\ (J_r \bullet B_2) - J_r & \text{if } r \text{ is even.} \end{cases} \tag{3}$$

The greyscale skeleton associates a vector $\mathbf{S}(x,y) = (S_1, S_2, \ldots, S_{15})(x,y)$ with each image pixel. This vector skeletonisation is converted to scalar skeletonisations as follows. First of all, noise reduction was applied by convolution along each vector $\mathbf{S}(x,y)$ with a one-dimensional Gaussian kernel, g_σ, with $\sigma = 2$ chosen as large as possible to retain the narrowest vessels. The result is $\mathbf{S}'(x,y) = \mathbf{S}(x,y) * g_2$. Then the maximum M is taken of the vector at each pixel;

$$M(x,y) = \max(\mathbf{S}'(x,y)). \tag{4}$$

The location of this maximum along the r-axis, to subpixel resolution, is denoted $R(x,y)$. It was taken as the location of the peak of a parabola fitted through the maximum of $\mathbf{S}'(x,y)$ and the two points in $\mathbf{S}'(x,y)$ either side of the maximum. See Figure 1(ii and iii).

A simplification of the skeleton is performed by keeping only ridges of M; the non-ridge pixels of M and the corresponding pixels of R are set to zero. A further simplification was obtained by removing "overlaps". These are pixels (x,y) for which there exists a pixel (x',y') with $M(x',y') \gg M(x,y)$ and with $R(x',y')$ greater than the distance from (x,y) to (x',y'). Evaluation of condition $M(x',y') \gg M(x,y)$ as $M(x',y') > 2\,M(x,y)$ gave good results.

A skeleton can be expressed as a graph in which nodes represent non-zero pixels and edges represent adjacency between these pixels. R and M can be expressed as the same graph, $\mathcal{G}$. The notation $R(n)$ and $M(n)$ will represent the values of R and M at node n of $\mathcal{G}$. Let $\mathrm{dist}_{\mathcal{G}}(n_1, n_2)$ be the distance between n_1 and n_2 defined as a count of the edges in

the shortest path through $\mathcal{G}$ from n_1 to n_2. Opening and closing of R with respect to $\mathcal{G}$ were performed to produce R_a (using erosion and dilation defined as the minimum and maximum over neighbourhoods with $\mathrm{dist}_{\mathcal{G}}(n_1, n_2) \leq 5$.) Weighted convolution was also performed with respect to $\mathcal{G}$ using

$$R_b(n) = \left[\sum_{d=0}^{d_{\max}} \sum_{\{m:\, \mathrm{dist}_{\mathcal{G}}(m,n)=d\}} g_d M(m) R_a(m) \right] / \left[\sum_{d=0}^{d_{\max}} \sum_{\{m:\, \mathrm{dist}_{\mathcal{G}}(m,n)=d\}} g_d M(m) \right] \tag{5}$$

where $g = (g_0, g_1, \ldots, g_{d_{\max}})$ is a Gaussian kernel, $g_d = (1/(\sigma\sqrt{2\pi}))e^{-d^2/(2\sigma^2)}$ with $\sigma = 5$. The weighting was provided by M since pixels in M with higher values correspond to more visually prominent features and so deserve higher weighting. The convolution, M_a, of M by g was calculated similarly but unweighted. Skeleton components due to image noise were removed by deriving a threshold T such that the pixels of M_a which are above T constitute 1.2% of the image area. This is well above the area that the skeletonised retinal vasculature is expected to cover. Nodes for which M_a is less than T were removed from $\mathcal{G}$ and the corresponding pixels in M_a and R_b were set to zero.

The image luminosity, L, at each pixel (x, y) on the skeleton was calculated as the negative mean intensity of I_S over a disc, $D^{(x,y)}_{R(x,y)}$, of radius $R(x, y)$ centred at (x, y);

$$L(x, y) = -\mathrm{mean}_{(x',y') \in D^{(x,y)}_{R(x,y)}} (I_S(x', y')) \tag{6}$$

To enable efficient application of the adaptive contrast evaluation and classification methods described in the next sections, the skeleton was broken into segments, $U_1 \ldots U_m$, referred to as vessel candidates.

2.3 Adaptive contrast evaluation

This section explains how a region for evaluation of local contrast was constructed around each vessel candidate. The watershed transform was applied to the gradient magnitude of $I_S * g_3$, where g_3 is a two-dimensional Gaussian kernel with $\sigma = 3$. Pixels on watershed boundaries were assigned to an adjacent watershed region. The watershed regions will be denoted W_k. The following is repeated for each vessel candidate, U_i. Let Q_0 be the set of watershed regions which intersect U_i. For $j = 1 \ldots 4$, define

$$Q_j = Q_{j-1} \cup \{W_k : W_k \text{ is adjacent to } Q_{j-1}\}. \tag{7}$$

Also define $A(W_k)$ as the lowest j such that $W_k \in Q_j$. $A(W_k)$ is an approximate measure of the distance of W_k from Q_0. The following processing is made more efficient by using only the smallest rectangular sub-image containing Q_4. A region growing method is performed with $P_0 = Q_0$ and $P_j = P_{j-1} \cup W_{k_j}$ where $W_{k_j} \in Q_4$ is a watershed region adjacent to P_{j-1} chosen to minimise

$$\left| \mathrm{mean}(I_S(W_{k_j})) - \mathrm{mean}(I_S(P_{j-1} - P_0)) \right| A(W_{k_j}) \tag{8}$$

For the case $j = 1$, the term, $\mathrm{mean}(I_S(P_{j-1} - P_0))$, cannot be evaluated and is replaced by 0. Region growing is stopped as soon as P_j covers at least half the area covered by Q_4, Figure 2 (i). The left hand term in (8) encourages choice of W_{k_j} with image intensity similar to that over P_{j-1}. The $A(W_{k_j})$ term encourages choice of W_{k_j} close to P_0. Image contrast was evaluated over a region $P_{\mathrm{con}} = P_j - P_0$, where P_j is the result of region growing. I_S was high pass filtered over P_{con} and the standard deviation (sd) was taken;

$$\mathrm{contrast}(U_i) = \mathrm{sd}_{(x,y) \in P_{\mathrm{con}}} \left(I_S(x, y) - \frac{\sum_{(i,j)} I_S(x+i, y+j) B_{\mathrm{con}}(x+i, y+j) g_3(i, j)}{\sum_{(i,j)} B_{\mathrm{con}}(x+i, y+j) g_3(i, j)} \right) \tag{9}$$

where $B_{\mathrm{con}}(x, y) = 1$ if $(x, y) \in P_{con}$ and is equal to zero otherwise and g_3 is a Gaussian, $\sigma = 3$.

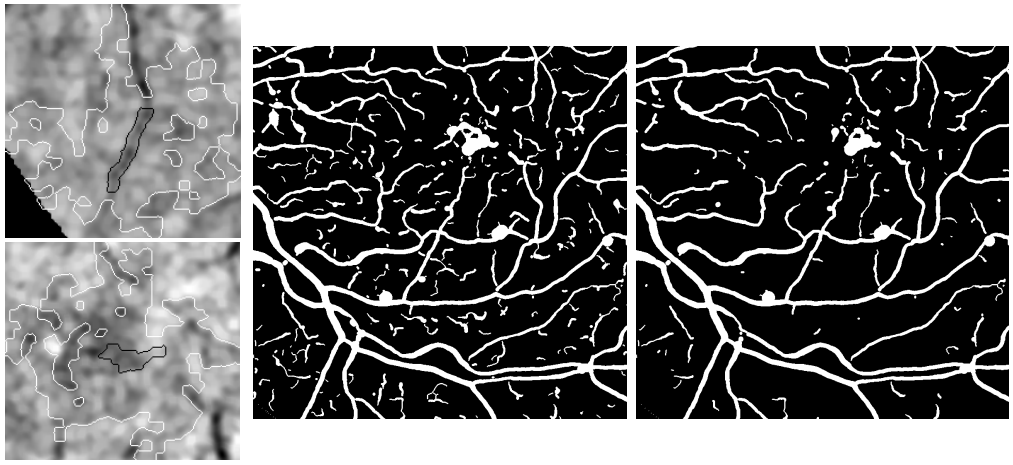
2.4 Classification

A kNN-classifier was used to classify candidate vessel segments U_i as follows. The features listed in Table 1 were calculated for each segment. The features were calculated for all candidate vessel segments with length greater than 20 pixels of 13 training images obtained from the DRIVE database [3]. The features were also calculated for candidate vessel segments in each test image. To equalise the significance of features in the classifier, training and test features were divided by the standard deviation of each feature over the training set. A kNN-classifier with $k = 25$ was applied to produce a soft classification of each candidate vessel segment which is equal to the proportion of the k nearest neighbours which correspond to true vessel segments in the training images.

Feature	Description	Feature	Description
$\bar{r}$	$\mathrm{mean}(R(U_i))$		
$\bar{m}$	$\mathrm{mean}(M(U_i))$	$\bar{m}_{\mathrm{norm}}$	$\bar{m}/\mathrm{contrast}(U_i)$
$\bar{l}$	$\mathrm{mean}(L(U_i))$	$\bar{l}_{\mathrm{norm}}$	$\bar{l}/\mathrm{contrast}(U_i)$

Table 1. Features calculated for each vessel segment, U_i.

A skeleton hard-classification, $V_{\mathrm{skel}}(t)$, is created at threshold $0 < t < 1$ of pixels in segments with soft classification greater than t. A non-skeleton hard-classification, $V(t)$, was created in which a pixel (x, y) was classified as vessel if there exists (x', y') whose distance from (x, y) is no more than $R_b(x', y')$ and if (x', y') is in $V_{\mathrm{skel}}(t)$.



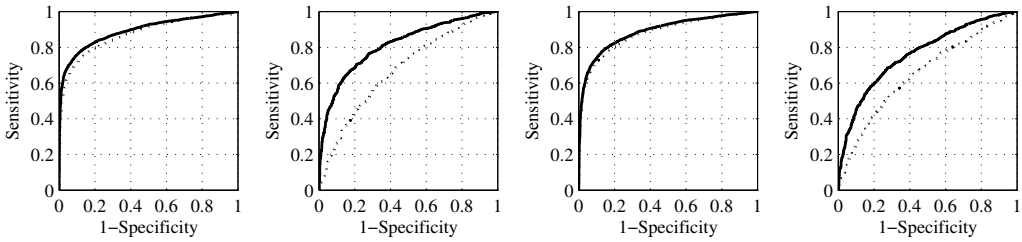
(i) (ii) (iii)

Figure 2. (i) Results of watershed region growing at vessel and non-vessel; black lines are the border of P_0, white lines are the border of P_j. Hard classifications; (ii) without contrast normalisation. (iii) with contrast normalisation.

3 Experiments

The methods were tested with the 20 test images in the DRIVE database [3]. Images were scaled to 1700 × 1600 (see above) and the final results, $V(t)$ and $V_{\mathrm{skel}}(t)$, were scaled back to the original image size. The performance with and without adaptive contrast enhancement was assessed. Classification was performed using feature sets $\{\bar{r}, \bar{m}, \bar{l}\}$ and $\{\bar{r}, \bar{m}, \bar{l}_{\mathrm{norm}}\}$. Pixel classification was assessed by comparing $V(t)$ to the manual segmentation that accompanies the database images. Vessel presence detection was assessed by comparing $V_{\mathrm{skel}}(t)$ with the manual segmentation, modified as follows. Narrow vessels in the manual segmentation were thickened by combining it with its skeleton, dilated with a disk of radius, $\rho = 1$, 2 or 3 pixels. This allows small location errors of narrow vessels to be tolerated.

4 Results



$A_z = \begin{cases} 0.8786\ (\bar{m}) \\ 0.8949\ (\bar{m}_{\mathrm{norm}}) \end{cases}$ (i) All

$A_z = \begin{cases} 0.6740\ (\bar{m}) \\ 0.8145\ (\bar{m}_{\mathrm{norm}}) \end{cases}$ (ii) Faint only

$A_z = \begin{cases} 0.8794\ (\bar{l}) \\ 0.8905\ (\bar{l}_{\mathrm{norm}}) \end{cases}$ (iii) All

$A_z = \begin{cases} 0.6534\ (\bar{l}) \\ 0.7633\ (\bar{l}_{\mathrm{norm}}) \end{cases}$ (iv) Faint only

Figure 3. ROC curves for classification of vessel and non-vessel segments by individual features (A_z is the area under the curves). Dotted lines are $\bar{m}$ and $\bar{l}$, solid lines are $\bar{m}_{\mathrm{norm}}$ and $\bar{l}_{\mathrm{norm}}$. (i) $\bar{m}$ and $\bar{m}_{\mathrm{norm}}$ all vessels. (ii) $\bar{m}$ and $\bar{m}_{\mathrm{norm}}$ faint vessels only. (iii) $\bar{l}$ and $\bar{l}_{\mathrm{norm}}$ all vessels. (iv) $\bar{l}$ and $\bar{l}_{\mathrm{norm}}$ faint vessels only.

Figure 3 compares the ability of individual features $\bar{m}$, $\bar{m}_{\mathrm{norm}}$, $\bar{l}$ and $\bar{l}_{\mathrm{norm}}$ to discriminate, in the training set images, between candidate vessels and non-candidate vessels using Receiver Operator Characteristic (ROC) curves. Figures 3 (i) and (iii) were produced using all 9,213 candidate vessel segments in the training images. There is only a very small improvement with adaptive contrast normalisation in these cases. Figures (ii) and (iv) were generated with 2,499 faint vessel candidates which satisfy the criterion $\bar{l} < 0.6$. These show a much greater difference between normalised and non-normalised features indicating that contrast normalisation is useful for distinguishing faint vessels from noise.

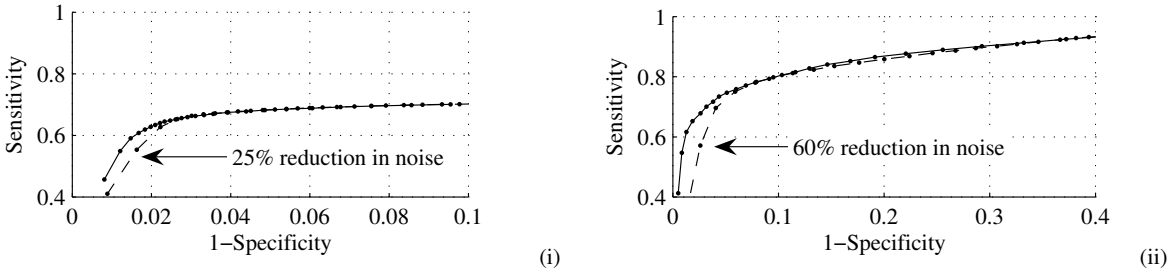


Figure 4. ROC curves for (i) pixel classification and (ii) vessel presence identification ($\rho = 2$).

Figure 2(ii) and (iii) shows example hard classifications without and with contrast normalisation. Figure 4 shows ROC curves for the classification of individual pixels and for identification of vessel presence. The improvement due to contrast normalisation is most pronounced at operating points with high specificity. Approximately 25% and 60% reduction in noise was achieved, due to contrast normalisation, with correct pixel identification and with correct vessel presence identification ($\rho = 2$, see Experiments), respectively. Similar improvements persisted if the tolerance to location errors was less (56% at $\rho = 1$) or more (70% at $\rho = 3$).

5 Discussion

This paper has shown that the recognition of the smaller and fainter retinal vessels is aided by adaptive contrast assessment. A method has been presented that allows a region of retina to be determined around vessel segments which does not intersect other nearby vessels or lesions making it suitable for adaptive assessment of local image contrast.

The greater difference between the curves in Figure 4(ii) compared to (i), corresponds with improvements in recognition of small vessels. Hence, this study suggests that incorrect recognition of intensity variations of the retina as vessels may be reduced by quantification of the local intensity variation. The same is true for other small dark features [1] and future work may show similar results for small bright features such as exudates, another feature of diabetic retinopathy. Certain applications require highly specific small vessel detection including vessel exclusion during lesion detection, fovea detection and assessment of small vessel visibility for checking image quality. Hence the improvements apparent in Figure 4 will assist this work to play an important role in an automated diabetic retinopathy detection system.

Acknowledgements

This work was funded by the Chief Scientist Office of the Scottish Executive Health Department.

References

1. A. D. Fleming, S. Philip, K. A. Goatman et al. “Automated assessment of diabetic retinal image quality based on clarity and field definition.” *Investigative Ophthalmology and Visual Science* **47(3)**, pp. 1120–1125, 2006.
2. A. Can, C. Stewart, B. Roysam et al. “A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24(3)**, pp. 347–364, March 2002.
3. J. Staal, M. D. Abramoff, M. Niemeijer et al. “Ridge-based vessel segmentation in color images of the retina.” *IEEE Transactions on Medical Imaging* **23(4)**, pp. 501–509, 2004.
4. X. Y. Jiang & D. Mojon. “Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images.” *IEEE Transactions on Pattern Recognition and Machine Intelligence* **25(1)**, pp. 131–137, 2003.
5. A. D. Fleming, S. Philip, K. A. Goatman et al. “Automated microaneurysm detection using local contrast normalization and local vessel detection.” *IEEE Transactions on Medical Imaging* In press.
6. D. Wu, M. Zhang, J. Liu et al. “On the adaptive detection of blood vessels in retinal images.” *IEEE Transactions on Biomedical Engineering* **53(2)**, pp. 341–343, February 2006.
7. S. Wang, A. Y. Wu & A. Rosenfeld. “Image approximation from gray scale medial axes.” *IEEE Transactions On Pattern Analysis And Machine Intelligence* **3(6)**, pp. 687–696, 1981.

MAX-MIN CENTRAL VEIN DETECTION IN RETINAL FUNDUS IMAGES

Hind Azegrouz [a] and Emanuele Trucco [a] *

[a]School of Engineering and Physical Sciences
Department of Electrical Electronic and Computer Engineering
Heriot Watt University, EH14 4AS Riccarton, UK

Abstract. This paper describes a new framework for the automated tracking of the central retinal vein in retinal images. The procedure first computes a binary image of the retinal vasculature, then obtains the skeleton (medial axis) of the vascular network. Terminal and branching points of the network are then located, and the network converted into a graph representation including length and thickness information for all vessels. Finally, a MaxMin approach is used to locate the central vein: candidate vessels are those associated to minimal paths from the optic disk to all terminal nodes found using Dijkstra algorithm. The actual central vein is selected among all the candidates by maximizing a merit function estimating the total vessel area in the image. Results are presented and compared with those provided by a manual classification on 20 images of the DRIVE set. An overall performance ratio of 92% is achieved.

1 Introduction

Central retinal vein occlusion (vein thrombosis) is a common pathology that affects the main central vein of the human retina, slowing the blood flow as it enters the optic nerve. Its causes are numerous and include glaucoma, hypertension, diabetes and poor ocular perfusion. Vein thrombosis can be manifested in other areas of the body and remains a serious disease affecting one in every 1,000 people in Britain [1]. Automatic or computer-assisted diagnosis of central retinal vein occlusion requires reliable and efficient detection and tracking of the central vein. This is the main objective of the work presented in this paper.

Many algorithms have been proposed for locating the vascular structure in retinal images [2–10]. Although some propose measures for vein/artery differentiation [11], vessel width [3, 12, 13] or vessel orientation [5, 14], or suggest a parabolic approximation of the main vessel [15, 16], no explicit method for the tracking of the central vein has been described. In this paper we present such a method using a max-min approach, and starting from a skeletonized version of the retinal vascular network. Our method relies on simple anatomical facts: the central vein is one of the longest retinal vessels, and veins are thicker than arteries.

In the rest of the paper a detailed explanation of the algorithm is presented, Results are shown and a conclusion is drawn.

2 Outline of the algorithm

The outline of the algorithm is presented in Figure 1.

*Email:ha19@hw.ac.uk, e.trucco@hw.ac.uk.

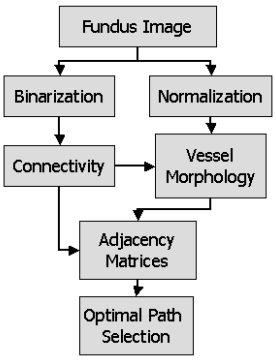
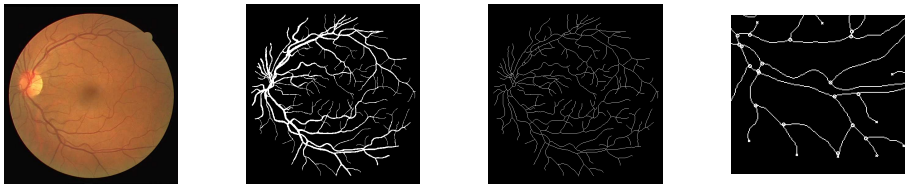


Fig. 1. Outline of the algorithm.

In a first step a binary image of the retinal fundus highlighting the vascular network is processed. We perform a skeletonisation [17], to obtain a thinned version of the previous image.The skeleton image is divided into 2 sub-images at the horizontal level of the optic disk. The graph-like structure of the 2 resulting images is exploited to draw a connectivity map and the location of terminal and branching nodes. For each arc in the graph, geometric properties such as length and thickness of the associated vessel are computed. The next module organizes all information obtained so far into adjacency matrixes. Dijkstra algorithm [18] is applied to find the minimal paths from the optic disk to each of the terminal nodes. These paths become the central vein candidates. The best candidate (and final result) is chosen by maximizing a merit function (hence MaxMin) estimating the overall area of the vessel associated to each candidate path.
The main steps of the algorithm are detailed in the following sections.

2.1 Obtaining the skeleton image

The input image is a binary map of the retinal vasculature (for example Fig. 2. (b)). We take this step for granted and use results of the publicly available DRIVE set by Staal et al [**?**]. We adopt MATLAB's skeletonisation algorithm based on iterative deletion of pixels, preserving the 8-neighbor connectivity information (Fig. 2. (a), (b), (c)).



(a) (b) (c) (d)

Fig. 2. (a) Fundus image of DRIVE set. (b) Segmented binary image of blood vessels. (c) The skeleton image. (d) example of terminal nodes and branching/bifurcation points

2.2 Detecting terminal and branching points

Using the skeleton image, two types of significant points can be detected: Terminal points and branching points. In order to classify a point as one, we count the number t of transitions from black to white while moving clock-wise around the 8-neighborhood of the point in question (Fig. 2. (d)).
t=1: determines a terminal node.
t=0, 2: determines a non-significant point.
$t \geq 3$: determines a branching point.

2.3 Connectivity of vertexes

The vascular skeleton is now interpreted as a graph, with terminal points and branching points as vertexes. The arcs are associated to portions of detected vessels.

Connectivity of vertexes, i.e., the existence of an arc linking them, is checked by deleting the vertexes and their 8-neighborhood, then labelling the connected components in the resulting image L; these are the edges of the graph, each of them will have a different label. Two vertexes i and j are then connected if their second order neighborhood V_i and V_j verify:

$L(V_i) \bigcap L(V_j) \neq \emptyset$

2.4 Vessel morphology

Once two vertexes i and j are proved to be connected, we compute the properties of the edge $E_{i,j} = [X_{i,j}, Y_{i,j}]$ linking them.

Length. The length of the arc is given by summing the Euclidean distances between each two neighboring points over the total n points of the arc.

$$L = \sum_{k=1}^{k=n-1} d([x_k, y_k], [x_{k+1}, y_{k+1}]) \tag{1}$$

Where d is the Euclidean distance.

Thickness. Thickness is computed as the median of vessel thickness values computed at each pixel in an arc. Thickness is computed by taking a cross-section of intensities perpendicularly to the vessel at a given pixel p_n, and fitting a Gaussian mixture model *G* (formed by 2 Gaussians) to the normalized intensity cross-section. The model is:

$$G = A + B\exp(\frac{-(z-\mu)^2}{2\sigma_1^2}) + C\exp(\frac{-(z-\mu)^2}{2\sigma_2^2}) \tag{2}$$

z being the coordinate along the local normal to the arc at each point p_n of the arc.

The fitting is an optimization problem in the space of the Gaussian model parameters, $[A, B, C, \mu, \sigma_1, \sigma_2]$. For the moment we adopt the Nelder-Mead simplex algorithm [19], but more efficient methods will be investigated in the future. The thickness of the arc is then chosen to be equal to $4\sigma_1$ knowing that $\sigma_1 > \sigma_2$. The median was chosen to exclude outliers, mainly due to poor model fit near vessel junctions (Fig. 3.).

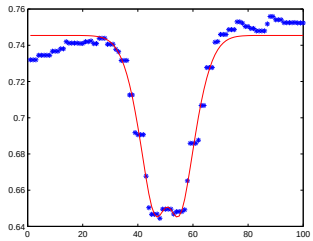


Fig. 3. Vessel cross-section

2.5 Building the adjacency matrixes

We number all the vertexes of the graph, and we build the following adjacency matrixes:

Connectivity matrix. The connectivity matrix $M_C$is such that each element $M_C(i,j)$ represents the connectivity between vertex i and vertex j. Thus:
$M_C(i,j) = 1$ if vertexes i and j are connected, i.e. belong to the same arc, and $i \neq j$.
$M_C(i,j) = \infty$ if i and j are not connected.
$M_C(i,j) = 0$ if i=j
$M_C(i,j) = M_C(j,i)$, as connectivity is a symmetric property.

Length matrix. The length matrix M_L is as follows:

$M_L(i,j) = L(i,j)$ if vertexes i and j are connected and $i \neq j$, $L(i,j)$ is the length of the edge between i and j
$M_L(i,j) = \infty$ if i and j are not connected.
$M_L(i,j) = 0$ if i=j.
$M_L(i,j) = M_L(j,i)$.

Thickness matrix. The thickness matrix is such that:

$M_T(i,j) = T(i,j)$ if vertexes i and j are connected and different, $T(i,j)$ is the thickness of the edge between i and j
$M_T(i,j) = \infty$ if i and j are not connected.
$M_T(i,j) = 0$ if i=j.
$M_T(i,j) = M_T(j,i)$.

2.6 MaxMin Central vein detection

As said, we partition each image into an upper and lower part at the horizontal level of the OD to make calculations faster. For each sub-image, we compute the 3 corresponding adjacency matrixes. From a vertex centered at the optic disk, we compute the *minimal* (shortest) path in term of the connectivity matrix M_T to each of the terminal nodes. The shortest paths are found using Dijkstra algorithm [18]. In the space of all minimal paths $[i_1, i_2, ..., i_{n_i}]$, where n_i is the number of arcs in the path, we *maximize* the following sum.

$$S = \sum_{k=1}^{n-1} M_T(i_k, i_{k+1})^{\alpha} M_L(i_k, i_{k+1}). \tag{3}$$

When $\alpha = 1$, S corresponds to the area of the vessel; when $\alpha = 2$, S is proportional to the volume of the vessel. We take $\alpha = 1.2$ as it achieves the best results in our experiments. The path corresponds to the retinal central vein, as the anatomy of the retina imposes that the central vein is one of the longest retinal vessels, and that veins are thicker than arteries.

3 Results

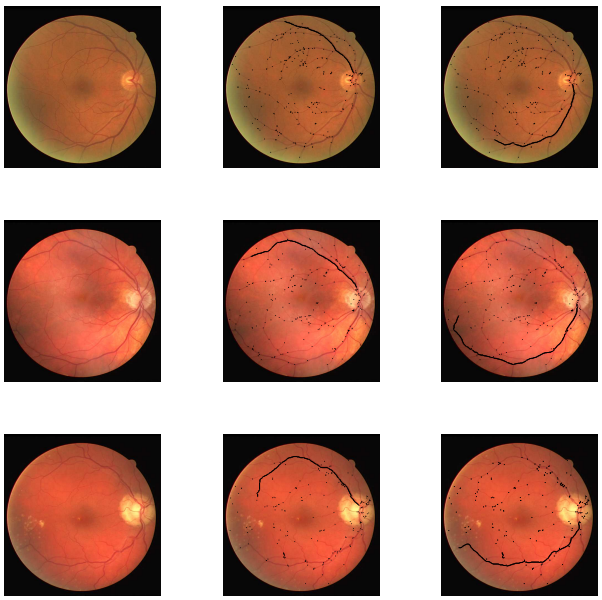


Fig. 4. Results showing the path of the central vein. The black points correspond to the vertexes of the vascular graph.

We ran the algorithm on a test set of 20 images from the DRIVE set [6]. A manual annotation of the images marking the path of the central vein and hence providing ground truth for comparison was previously done by an expert. To determine the classification error, we define the success rate as the number of pixels correctly classified (given the ground truth) as belonging to the main vessel divided by the total number of points of the actual central vein. An

average success rate (true positives) of 92% was obtained for the images in the test set (percentage of pixels correctly classified as on the target vein). Most errors are concentrated in the peripheral part of the path, where the resolution of the image limits the performance of the algorithm.

4 Conclusion

In this paper we have presented a new automatic method to track the central vein from fundus retinal images. The method uses connectivity and anatomy information to determine the minimal path in the vessel graph that maximizes a cost function. For the cost function described here we used only length, connectivity and thickness of the graph arcs. Better results are to be expected when branching angle values are included into the cost function, which is one of our current objectives. In parallel we are developing an algorithm to discriminate between veins and arteries, which should improve the performance of the vessel tracking even more.

References

1. C. Tovey & S. Wyatt. "Diagnosis, investigation, and management of deep vein thrombosis." *British Medical Journal* **326**, pp. 1180–1184, 2003.
2. O. Chutatape, Z. Liu & S. M. Krishnan. "Retinal blood vessel detection and tracking by matched gaussian and kalman filters." *Proc. Eng. in Medicine and Biology Society* **6**, pp. 3144 – 3149, 1998.
3. L. Gang, O. Chutatape & S. M. Krishnan. "Detection and measurement of retinal vessels in fundus images using amplitude modified second-order gaussian filter." *IEEE Trans. Biomedical Engineering* **49**, pp. 168–172, 2002.
4. A. Hoover, V. Kouznetsova & M. Goldbaum. "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response." *IEEE Trans. Medical Imaging* **19(3)**, pp. 203 – 210, 2000.
5. F. Zana & J.-C. Klein. "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation." *IEEE Trans. Image Processing* **10(7)**, pp. 1010–1019, 1998.
6. J. J. Staal, M. D. Abramoff, M. Niemeijer et al. "Ridge based vessel segmentation in color images of the retina." *IEEE Trans. Medical Imaging* **23**, pp. 501–509, 2004.
7. R. Polli & G. Valli. "An algorithm for real-time vessel enhancement and detection." *Comput. Meth. Programs. Biomed* **52**, pp. 1–22, 1997.
8. T. C. Henderson & G. Choikim. "An algorithm for real-time vessel enhancement and detection." *Proceedings. 7th Int. Symp. Signal Processing and Its App.* **1**, pp. 685 – 688, 2003.
9. Y. Sun. "Automated identification of vessel contours in coronary arteriograms by an adaptive tracking algorithm." *IEEE Trans. Medical Imaging* **8**, pp. 78–88, 1989.
10. Y. Tolias & S. Panas. "A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustring." *IEEE Trans. Medical Imaging* **17**, pp. 263–273, 1998.
11. E. Grisan & A. Ruggeri. "A divide et impera strategy for automatic classification of retinal vessels into arteries and veins." *EMBS* pp. 890–893, 2003.
12. L. Pedersen, M. Grunkin, B. Ersboll et al. "Quantitative measurement of changes in retinal vessel diameter in ocular fundus images." *Pattern Recognition Letters* **21**, pp. 1215–1223, 2000.
13. R. Bernardes, J. Dias & J. Cunha-Vaz. "Mapping the human blood-retinal barrier function." *IEEE Trans. Biomedical Engineering* **52**, pp. 106–116, 2005.
14. M. E. Martinez-Perez, A. D. Hughes, A. V. Stanton et al. "Retinal vascular tree morphology:a semi-automatic quantification." *TBE* **49**, pp. 912–917, 2002.
15. E. Trucco & P. J. Kamat. "Locating the optic disk in images via plausible detection and constraint satisfaction." **1**, pp. 155–158, 2004.
16. H. Li & O. Chutatape. "Automated feature extraction in color retinal images by a model based approach." In *IEEE Trans. Biomedical Engineering*, volume 51, pp. 246 – 254. 2004.
17. L. Lam, S. W. Lee & C. Suen. "Thinning methodologies-a comprehensive survey." *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**, pp. 879, 1992.
18. A. S. Tanenbaum. *Computer Networks*. Pearson Education International, 2003.
19. J. A. Nelder & R. Mead. "A simplex method for function minimization." *Computer Journal* **7**, pp. 308–313, 1965.

Web Services for Digital Mammography Research

Chris Rose, Daniele Turi, Alan R Williams, Katy Wolstencroft and Chris Taylor

Imaging Science and Biomedical Engineering,
The University of Manchester,
Manchester, M13 9PT, United Kingdom
http://www.medicine.manchester.ac.uk/isbe/

Abstract. The Digital Database for Screening Mammography (DDSM) is an invaluable resource for digital mammography research. However, there are two particular shortcomings that can pose a significant barrier to many of those who may want to use the resource: 1) the actual mammographic image data is encoded using a non-standard lossless variant of the JPEG image format; 2) although detailed metadata is provided, it is not in a form that permits it to be searched, manipulated or reasoned over by standard tools. This paper describes web services that will allow both humans and computers to query for, and obtain, mammograms from the DDSM in a standard and well-supported image file format. Further, this paper describes how these and other services can be used within grid-based workflows, allowing digital mammography researchers to make use of distributed computing facilities.

1 Background

The DDSM [1] provides high-resolution digitised mammograms, ground-truth (abnormality annotations provided by expert mammography radiologists) and metadata (including the date of study and digitisation, the Breast Imaging Reporting and Data System (BI-RADS) [2] breast density and assessment categories, a subtlety rating, the type of pathology and detailed categorisation of the nature of the perceived abnormality using the BI-RADS lexicon). The DDSM is available free of charge by File Transfer Protocol (FTP).

While the DDSM does provide software to decode their mammograms[1], the default distribution of this software does not build under modern compilers without modification, a step that may prove difficult to those with insufficient experience of C/C++ software development for UNIX-like operating systems. Furthermore, even when properly compiled, the DDSM software outputs the image data as a stream of raw bytes; one then has to normalise these according to the model of digitiser used to image the original films and then create an image file that is readable by one's image analysis software environment. An introduction to web services is given in Section 1.1. Section 2.1 describes a web service that allows digital mammograms from the DDSM to be obtained in a standardised and well-supported lossless file format. Section 2.2 describes a service that allows groundtruth images to be obtained in the same file format.

While a web-based query tool is provided by the DDSM, it is useful only to human users or automated tools that have been specifically designed with the DDSM in mind. If the metadata were in a more useful format, it could easily be exposed for both human and computer use. Section 2.3 describes a formal ontology that has been developed to describe the DDSM resource and a web-based user interface to allow users to query the ontology.

Section 2.4 describes how web services can be used together within *workflows* to run full experiments and how a full record of how such experiments were performed can be recorded by capturing provenance events. Section 2.5 details a supporting website for the work described in this paper.

1.1 An Introduction to Web Services

The concept of web services may best be explained with a simple example of a hypothetical scientist named Bob who lives in Manchester, UK. Bob has a computer-aided detection/diagnosis (CAD) algorithm that other scientists want to use. Traditionally, Bob would package his CAD algorithm into some form that is easily installed and run by other scientists. He would then deliver it to those scientists via Internet download or on physical media (e.g. CD-ROM). However, Bob might not be able to let other scientists run his software on their computers because: Bob may not have planned to share his software and may have made assumptions in its design that limits its portability; users might need an expensive license to use a required proprietary library; the software may need access to a resource (e.g. a large database) that resides at Bob's lab; Bob might frequently update his software, so making each update available to all his users might be troublesome; packaging the software for easy installation might be too time-consuming for a busy research scientist.

[1] In particular the DDSM's `jpeg` program.

Bob might decide that it is easier to allow other scientists to run his software on his computer or a server, allowing access to it via the Internet. This can be achieved by exposing his software as a web service. Using an appropriate piece of client software that implements the Simple Object Access Protocol (SOAP) standard [3], other scientists can run Bob's CAD algorithm on their data. In this way, Bob's CAD algorithm can be used by remote scientists as if it were installed on their computers, or integrated into software as if it is a library containing the required functionality.

The "interface" to a web service—the location of the computer that provides the service and a specification of its inputs and output and their data types—is described using the Web Service Description Language (WSDL) [4]. The URL of a service's WSDL file is all that is needed for a SOAP client to be able to use the service[2].

The web services proposed in this paper open up the possibility of performing digital mammography research using grid-based workflows. In this context, a *grid* is an *ad hoc* collection of computing services offered by a number of (typically) different computers via a network (typically the Internet) and a *workflow* specifies how the services provided by the grid are orchestrated in order for some task to be performed. Section 2.4 describes workflows in more detail.

2 Method

2.1 Digital Mammogram Web Service

To create a web service that makes available mammograms from the DDSM, we developed a command-line program that, given the name of a particular DDSM mammogram (e.g. D_0160_1.RIGHT_MLO), downloads the associated `.LJPEG` file from the DDSM's FTP server, decodes the raw image data, normalises it according to the digtiser used and finally converts it to a PNG file [5]. This format is ideal for encoding mammograms as it is standardised, guarantees lossless compression and is widely supported by common software tools and libraries[3]. (In future, other lossless image formats may be more suitable—such as JPEG-LS and JPEG 2000—but as of this writing these formats are not widely supported.)

Downloading and converting the images takes a few minutes on a desktop computer with a fast connection to the Internet and so our service currently caches some of the converted mammograms locally so that future requests can be efficiently serviced[4]. Our program is exposed as a service via SOAP [3]. DDSM mammograms can be obtained from this service using any SOAP client.

2.2 Groundtruth Web Service

To enable the evaluation of algorithms run on the DDSM mammograms delivered by the service described in Section 2.1, there is a need to be able to obtain the corresponding groundtruth images. We have developed a web service that allows DDSM groundtruth images to be generated and delivered in a suitable image format. Our approach to developing this service was the same as that described in Section 2.1. We first developed a command-line program that, given the name of a particular DDSM mammogram (e.g. D_0160_1.RIGHT_MLO) and an abnormality number[5], downloads the corresponding `.OVERLAY` file (which contains a description of the shape of the radiologist-annotated abnormalities for the image using a chain code). The DDSM groundtruth metadata defines two possible types of region: a *'boundary'* and a *'core'*, though *'core'* regions may be absent (see [1] for details). Our program then creates a image with the same number of rows and columns as the corresponding digital mammogram—i.e. there is a one-one correspondence between every pixel in a digital mammogram and the pixel at the same location in its groundtruth image—and this image is populated with pixel values that indicate the class of each pixel. A value of zero represents normal tissue or the non-breast region, a value of 128 represents a pixel within a *'boundary'* region of the abnormality and a value of 255 represents a pixel within a *'core'* region of the abnormality. The resulting groundtruth image is saved as a PNG image. While downloading the `.OVERLAY` file and creating a groundtruth image takes approximately a minute, the resulting PNG files are very small (approximately 14 kb) due to the high correlation between successive scan lines in the images, and we therefore keep all generated images to efficiently service future requests. This program is then exposed as a web service via SOAP.

[2]The WSDL files for the services described in this paper can be obtained from the website described in Section 2.5

[3]We encode DDSM mammograms as 16 bits/pixel grey-level images.

[4]We are in the process of implementing a hardware and software infrastructure to allow all mammograms to be stored locally, which will improve the speed with which requests can be services substantially.

[5]Mammograms may contain more than one abnormality. The abnormality number is captured by the ontology that is described in Section 2.3

2.3 DDSM Ontology and Metadata Query Service

Being able to obtain mammograms and groundtruth is not particularly useful without knowing which mammograms have what characteristics. To this end, we have developed a formal *ontology* of the mammograms, groundtruth and metadata (e.g. abnormalities, patient information). An ontology is a description of the concepts and relationships that exist in some knowledge domain. The formal representation of metadata within ontologies (using technologies such as the Web Ontology Language (OWL) [6]) allows domain knowledge to be used alongside explicit labelling to infer implicit relationships and hence deliver more useful results. As a simple example, if an ontology were to state that a stellate lesion is a type of mass, then a query for masses could return—in addition to items explicitly labelled as masses—items that were labelled as stellate lesions; i.e. the domain knowledge captured in the formal ontology allows it to be inferred that items labelled as stellate lesions must also be returned.

2.3.1 Our DDSM ontology

Within the DDSM, information about the mammograms is specified in an `.ICS` file and, for each image that contains abnormalities, an `.OVERLAY` file. The `.ICS` file contains information common to the case, e.g. the patient's age, and also information necessary to interpret the four mammograms, e.g. the number of pixels per scan line. The `.OVERLAY` file contains information particular to the abnormality, or abnormalities, that have been interpreted within a particular mammogram, e.g. the left CC mammogram.

The ontology that has been developed for the DDSM allows the description of the information within the `.ICS` and the `.OVERLAY` files. The ontology is written in OWL [6]. A decision was made that the ontology would only describe the information specified within the DDSM, in particular it would not attempt to be a general model of mammograms, mammogram interpretation or breast cancer as we are interested only in making the DDSM database easily available. However, the apparently limited scope of our ontology allows for greater flexibility: if a user requires an ontology that models a wider domain our ontology can be mapped onto the user's ontology. This allows our ontology to be used in wider contexts without requiring us to (potentially incorrectly) anticipate what other users will want.

For an individual case, an OWL ontology is populated with RDF (Resource Description Framework [7]) instances. The instance ontology combines the information within the `.ICS` and `.OVERLAY` files into a single semantic structure. This allows the easy searching of the instance ontology when it is loaded into an RDF repository.

Within the DDSM ontology, the *'case'* class specifies information that applies to a patient's visit and their four images. It has four relationships to *'views'* corresponding to the four mammograms. *'Views'* are subclassed into either abnormal or normal views. The information about the image is held within the *'view'* superclass. If the *'view'* is an *'abnormal view'* then it has relationships to one or more *'abnormalities'*.

An *'abnormality'* contains information such as the assessment and subtlety. It also has specific information about the calcification or mass intrepretation of the abnormality. In addition, the bounding curves of the abnormality and any cores within it are specified.

We have automatically populated an RDF store with instances of the classes in our ontology by processing the DDSM metadata files. We are currently developing a web-based user interface that will allow users to query the RDF store for images and groundtruth in a user-friendly manner. Figure 1 shows the form used to create queries.

2.4 Workflow Enactment and Provenance Capture

A workflow describes how a number of services can be combined to perform some useful task. The Taverna workbench program—a Java application that originated in the bioinformatics research community—allows users to create and run workflows within a graphical user interface [8]. Taverna displays workflows as directed graphs, where nodes represent inputs, services or outputs and arcs represent the flow of data and control (see Figure 2 for an example). Workflows can be saved and easily exchanged between colleagues. Taverna allows users to run, pause, monitor and debug workflows in a manner similar to modern software development environments. Workflow results can be directly displayed within Taverna.

Aside from allowing researchers to make use of distributed computing resources, Taverna can capture *provenance events*—e.g. when a particular workflow was started and with which inputs—allowing the workbench to operate as an automated laboratory log book. This also allows researchers to obtain answers to questions like *'which images were*

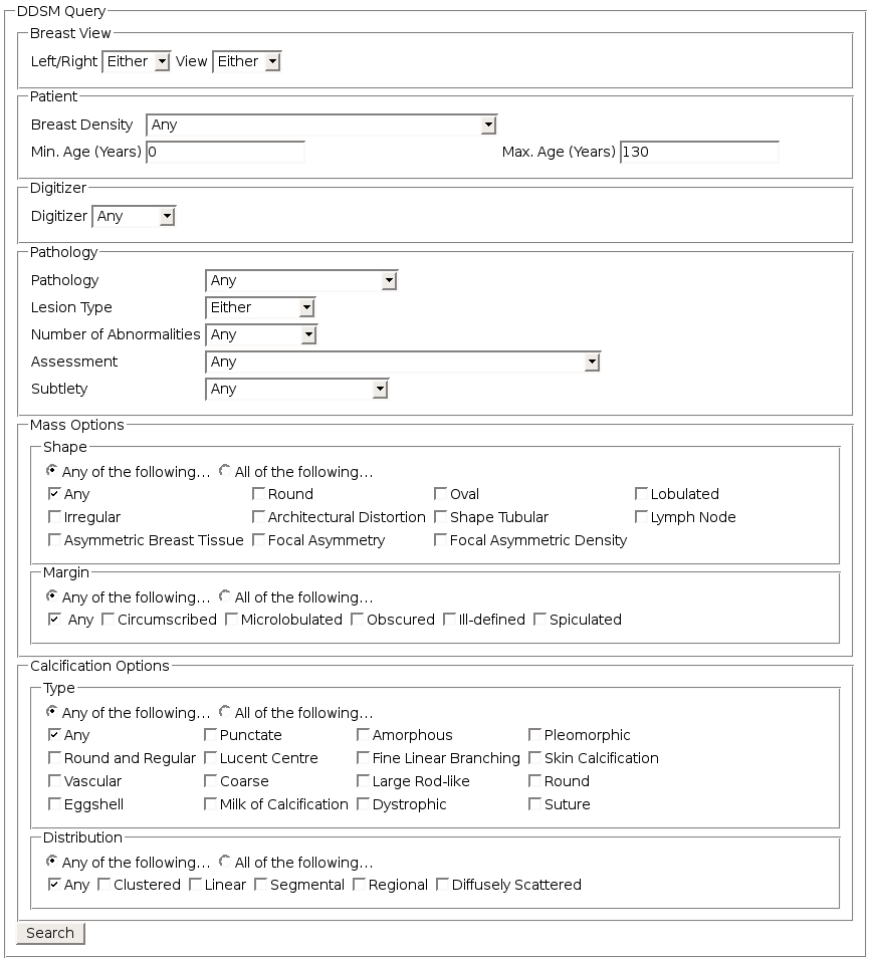


Figure 1. The DDSM metadata query form.

used as inputs to the workflow I ran on 24 January 2006?'.

2.5 Supporting Website

Those who wish to use the services described in this paper are directed to [9]. This website will provide the most up-to-date documentation of the available services and provide links to the WSDL files that specify the web services described in this paper. We invite the community to make useful software and data available via services and are keen to document these at the above website.

3 Results

The most significant contributions are the digital mammogram and groundtruth "getter" services, the DDSM ontology, the web-based query facility and the supporting website. We hope these will be useful to the digital mammography research community. While we are using the Taverna workbench software to integrate our services into grid-based workflows, our general approach—publishing our software as SOAP services—does not require clients to use Taverna; any SOAP client can be used.

4 Discussion

We have described three web services which allow both humans and computers to query a formal ontology of the DDSM data and obtain digital mammograms and groundtruth from the DDSM in a well-supported standard image format. We have also described how these services could be used within grid-based workflows. As Section 1 described, obtaining mammograms from the DDSM is currently non-trivial and it is hoped that the web services described in this paper will make using this important resource more convenient.

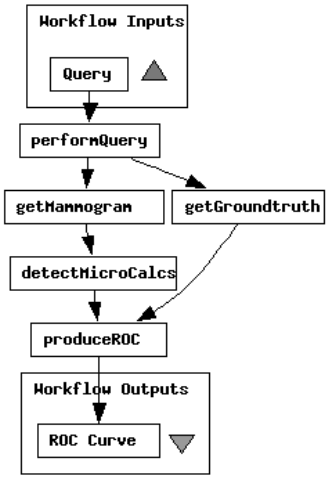


Figure 2. A simple Taverna workflow.

It is difficult to quantitatively evaluate the type of work that is described in this paper. While we could measure the speed with which requests can be processed, or the number that can be handled concurrently, such measurements do little to tell us if we have achieved our aims of developing and deploying infrastructure that is useful to the community. This will become apparent in time as the resources described in this paper are used (or otherwise) and if other researchers contribute their software and data in the form of web services for use by the community. In this spirit we welcome criticism and suggestions and are able to offer advice on an informal basis to those interested in developing their own web services.

Future work will focus on exposing other useful algorithms as web services (e.g. a CAD task such as microcalcification detection, a receiver operating characteristic (ROC) analysis service) and on maintaining the website described in Section 2.5. Given these services, it will be possible to run a simple but typical digital mammography CAD experiment using web services (i.e. obtain mammograms → process each image using the CAD algorithm → obtain ground-truth → produce an ROC curve). By publishing the workflow, others in the community would be able to replicate the experiment exactly or swap one service (e.g. the CAD task) for their own to be able to fairly compare algorithms.

References

1. M. Heath, K. Bowyer, D. Kopans et al. "The Digital Database for Screening Mammography." In M. J. Yaffe (editor), *Digital Mammography: IWDM 2000, 5th International Workshop*, pp. 212–218. Medical Physics Publishing, Madison, Wisconsin, USA, December 2001.
2. "The ACR Breast Imaging Reporting and Data System (BI-RADS)." American College Radiology, 1998. Third Edition.
3. M. Gudgin, M. Hadley, N. Mendelsohn et al. "SOAP Version 1.2 Parts 1–2." W3C Recommendation, World Wide Web Consortium, June 2003.
4. E. Christensen, F. Curbera, G. Meredith et al. "Web Services Description Language (WSDL) 1.1." W3C Note, World Wide Web Consortium, March 2001. (A W3C Recommendation for WSDL 2.0 is currently pending.).
5. D. Duce. "Portable Network Graphics (PNG) Specification (Second Edition)." W3C Recommendation, World Wide Web Consortium, November 2004.
6. D. L. McGuinness & F. van Harmelen. "OWL Web Ontology Language Overview." W3C Recommendation, World Wide Web Consortium, February 2004.
7. F. Manola & E. Miller. "RDF Primer." W3C Recommendation, World Wide Web Consortium, February 2004.
8. T. Oinn, T. Greenwood, M. Addis et al. "Taverna: Lessons in creating a workflow environment for the life sciences." *Concurrency and Computation: Practice and Experience Grid Workflow Special Issue (Accepted)* 2005.
9. "Digital Mammography Services Network." `http://www.digital-mammography-services.net`.

Cervical Cancer Detection from Phase Contrast Cytological Images

Marcin Smereka[a*] and Grzegorz Glab[b]

[a]Institute of Computer Engineering, Control & Robotics, Wroclaw University of Technology,
Janiszewski Str. 11/17, 50–372 Wroclaw, Poland,
[b]Gynaecological Clinic GMW, 1st Maj Str. 9/72, Opole, Poland

Abstract. In this paper a new approach to cervical cancer detection is presented. It is based on an automated analysis of cytological images acquired by the phase contrast microscope. This technique is an alternative to traditional Pap smear tests. The paper presents a practical combination of image processing and pattern recognition algorithms aimed to identify pathological and atypical cells. All the exploited algorithms were customized to reflect the specificity of phase contrast images and apriori–knowledge of cytological smears.

1 Introduction

The diagnostic cytology is an integral component of gynaecological examinations. It enables early detection of precancerous lesions in the uterine cervix. The phase contrast (Ph) microscopy is relatively a new technology in this area. Traditional Pap–Smear tests require staining and fixing, what usually takes several days before the diagnosis is issued. Using Ph microscope, a non–fixed and non–coloured cellular smear can be observed just after placing it on a micro slide [17], therefore it allows for immediate diagnosis [1]. Visualization of intracellular structures is possible due to diffraction and interference of light waves around different morphotic elements (Fig. 1a). Natural biological preparations (cells, tissues) consist mostly of a great number of irregularly distributed elements. These elements are optically heterogeneous, thus their refractivity differs and the phase of light waves is shifted. Because both direct and diffracted light can not be separated completely, optical artifacts appear in images. In literature [5] the artifacts are known as 'halo' and 'shading off' effects. These effects tend to emphasize edges and boundaries of observed elements (Fig. 1b). Computer aided processing of Ph images can additionally improve performance and quality of examination. The joint research of authors is aimed at recognition and classification of objects present in the images. It is our belief that computer image processing system could effectively perform the preliminary screening, and provide only pathological and atypical cells for physician's evaluation. Interesting medical objects occur rarely in early cancer phases, so the system could prevent from monotonous inspecting of hundreds of microscopic images by a human. Another benefit of using this technique relies on easiness of coupling cytology and colposcopy examinations, what improves reliability of cervical diagnostics [2]. A carcinogenesis on uterine cervix is a multi–step and long–lasting process. Through analysis of the specified features of the squamous cells the smear can be classified as normal or pathological one (ASCUS, L–SIL, H–SIL).

2 Related work

Many automated screening systems have been developed for stained and fixed cytological smears. Techniques are mostly based on colour and texture information, therefore they are not applicable to Ph images. In our previous work some algorithms based on geometric features have been tested (the Hough Transform, shape descriptors, active contours) to detect shapes from Ph images. The Hough Transform (HT) has been recognized as a very powerful method

*Email: Marcin.Smereka@pwr.wroc.pl

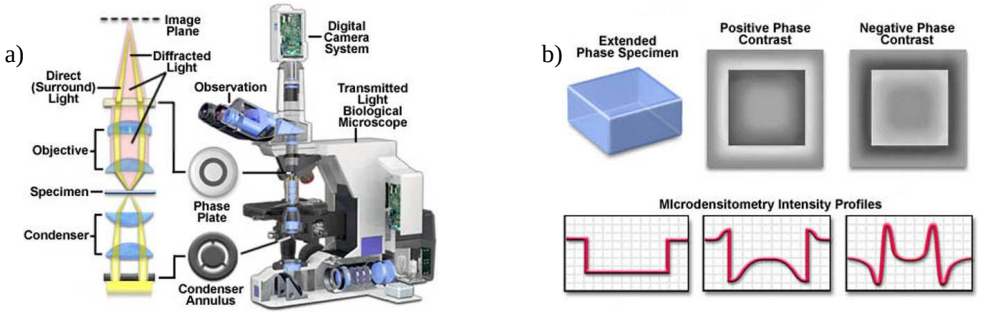


Figure 1. a) Ph microscope; b) 'Halo' and 'shading off' in positive and negative Ph. Source: [17]

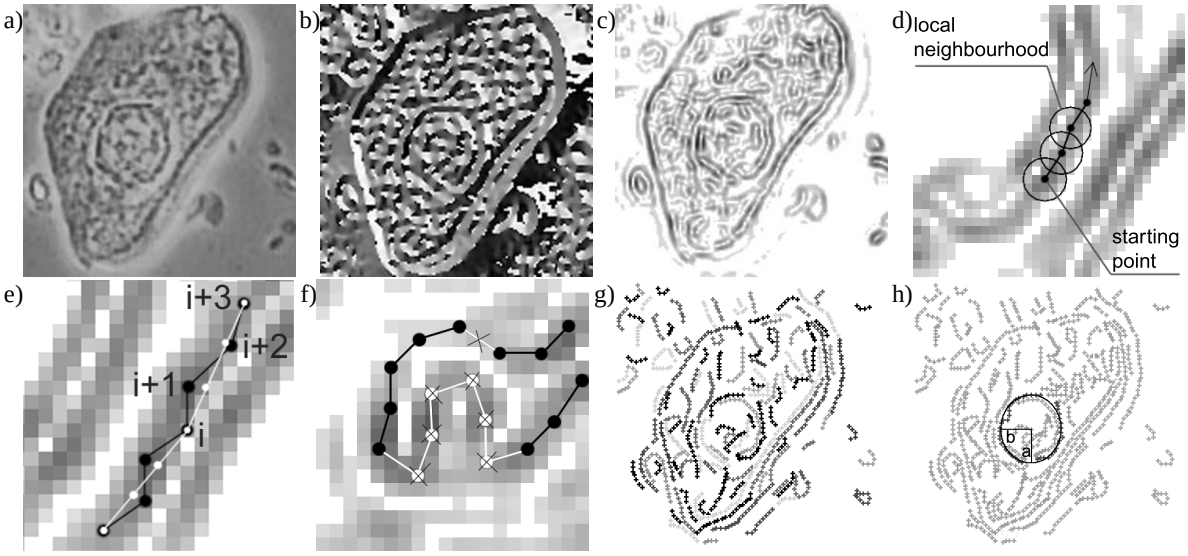


Figure 2. Example of Ph image processing. a) original image, b) orientation image encoded by graylevels I_φ (black — $0°$, white — $359°$), c) amplitude image I_A, d) a ridge following, e) 'staircase' and smoothing, f) a segment after removing highly curved points and inflexion points, g) an image processed by the ridge following algorithm, h) an instance of the path

to detect parametric curves in images [8]. The direct HT method for detecting ellipses is computationally expensive due to the multidimensional parameter space. Many improvements have been proposed to make these methods more efficient and more robust to irregularities. Improved HT was successfully applied to detect regular cell nuclei from Ph cytological images [3], but it hardly dealt with pathological cells. Irregular shape of pathological nuclei provided relatively weak peaks in a parameter space of HT. A contour grouping algorithm was successfully used in [4] to isolate irregular shape nuclei. The present work combines the idea of contour grouping with pattern recognition methods in order to detect cell nuclei and provide diagnostic information.

3 Method

3.1 Acquisition and Initial filtering

The computer system controls a mobile platform of microscope and acquire digital images by a frame grabber connected to CCD camera. Using a mobile platform the system takes series of adjacent images of the smear and processes them separately. An acquired image should be initially filtered in order to fulfil two goals. Firstly, to ignore chrominance (colour information), i.e. to convert image from RGB to gray scale. Note, that in case of Ph microscopy, only the luminance carries useful information about visible objects. To preserve required contrast in the image, a gamma correction is used and coefficients of R,G,B components are applied. Secondly, to equalize the background level (if the illumination of the smear is not uniform, a feature extraction is more difficult).

3.2 Segmentation and feature extraction

It is assumed that the most significant objects in Ph images are cell nuclei. The size and the shape of a nucleus brings a lot of information about precancerous lesions in progress. Large, irregularly outlined and non-uniformly filled nuclei are suspected to be atypical (ASCUS) or pathological ones (L–SIL, H–SIL) [1]. The goal of segmentation algorithm is to detect cell nuclei and process them afterwards. The algorithm is based on the following assumptions: the Ph microscopy emphasizes edges of objects [5], thus the nuclei detection is equivalent to searching of their boundaries; the shape of nucleus is circular or elliptic, so oval patterns are of particular interest; the image magnification is known in advance, thus only the objects within the specific range of radii $[r_{min} = 10, r_{max} = 35]$ are considered. The range was fixed based on hundreds of samples. The presented segmentation algorithm follows three–step procedure. At first, edges and their orientation are detected. Then, edges with high curvature are removed. Finally, edge grouping is performed and features are extracted for each group of edges.

The goal of edge detection is to transform initial gray scale image (Fig. 2a) to an edge image which assigns amplitude and orientation of edge gradient to each point. At first, the initial image is smoothed with Gaussian kernel, then it is convoluted with horizontal and vertical Canny operators. As the result, horizontal and vertical gradient images are obtained. Prewitt and Sobel gradient operators also can be applied for this purpose. Zero–crossing detectors and morphological methods failed easily because they generated too many false edges. To compute orientation image I_φ and amplitude image I_A (Fig. 2b,c) Cartesian coordinates are transformed to polar ones.

A cytoplasm of a cell in Ph image consists of randomly placed spots and highly curved short line segments. These shapes can be easily misrecognized as nuclear walls or cell walls. To separate real and phantom edges an algorithm was developed to extract long and smooth edges. This algorithm finds the initial point (the point with the maximum amplitude), follows the edge in both directions using edge orientation (I_φ) and stops eventually, where the edge amplitude drops below a given threshold t_A. To avoid side effects such as loops and adjacent ridges, it is assumed that the ridge is $w = 5$ pixels wide and the local neighbourhood of a pixel is removed from image after the pixel selection (Fig. 2d). The value of w corresponds to the size of edge operators and defines the minimum distance between adjacent ridges. When w is too small the ridge can turn back along the same edge, while too large value of w can make the ridge jump to the neighbouring or adjacent edge. Pixels that form one ridge are removed from the image and the process is repeated until no new starting point with amplitude larger than t_A is available. The choice of the optimal value of t_A is a trade–off between sensitivity for weak and soft edges and performance of the grouping process. Experiments shown that the best results were obtained for $t_A = 0.25 \max I_A^{(x,y)}$. After all segments were created, each segment must be freed of the 'staircase' effect (Fig. 2e). To accomplish this, each quadruple of consecutive points (i,i+1,i+2,i+3) for each edge segment is tested. In the case of inflexion ($infl < 0$) internal pixels (i+1,i+2) are smoothed linearly.

$$\Delta x_i = x_{i+1} - x_i; \quad \Delta y_i = y_{i+1} - y_i \tag{1}$$

$$infl = (\Delta x_{i+1} \cdot \Delta y_i - \Delta x_i \cdot \Delta y_{i+1}) \cdot (\Delta x_{i+2} \cdot \Delta y_{i+1} - \Delta x_{i+1} \cdot \Delta y_{i+2}) \tag{2}$$

After this operation, coordinates of points have subpixel accuracy. Then the curvature κ_i is calculated for each point.

$$\Delta\tilde{x}_i = \frac{\Delta x_i}{\sqrt{\Delta x_i^2 + \Delta y_i^2}}; \quad \Delta\tilde{y}_i = \frac{\Delta y_i}{\sqrt{\Delta x_i^2 + \Delta y_i^2}}; \quad \kappa_i = (\Delta\tilde{x}_i - \Delta\tilde{x}_{i-1})^2 + (\Delta\tilde{y}_i - \Delta\tilde{y}_{i-1})^2. \tag{3}$$

Points with curvature over the threshold value t_κ are removed and segments are split into smaller parts (Fig. 2f). The value $t_\kappa = 2$ was chosen experimentally and corresponds to the angle $\pi/2$. Additionally, inflexion points and very short segments are removed in order to speed up further processing. As the result of ridge following, a set of edge segments is obtained (Fig. 2g).

Let us define a graph G, where vertices correspond to edge segments and arcs connect pairs of segments placed close enough to each other. Two segments are close to each other if the distance between their centroids is smaller than $2r_{max}$. A boundary of a nucleus consists of one or more segments. The goal of contour grouping algorithm is to group segments that form the nucleus boundary, so searching for nucleus is equivalent to searching for a path in the graph. The search algorithm extracts all possible paths in G containing up to 4 vertices.

For each tested path the ellipse–fitting algorithm [6] is executed to form the best–fit ellipse to the edge points of the path(Fig. 2h). If the ellipse meets the size constrains: $r_{min} \leq a \leq r_{max}$ and $b/a > 0.5$ numeric features collected in Table 1 are calculated for the path. Some features were selected to distinguish real object from phantom objects(random set of edges), when the other features were designed to detect pathologies. Symbols $\searrow$, $\nearrow$ denote low/high level, respectively.

Name	Interpretation
length of the major semi–axis of ellipse (a)	$\searrow$ probably regular, $\nearrow$ probably pathological
avg. amplitude on the boundary of ellipse	$\searrow$ probably background, $\nearrow$ sharply outlined object
avg. angle between edge orientation and corresponding ellipse orientation	$\searrow$ elliptic object, probably nucleus, $\nearrow$ random edges
error of fit between ellipse and edge segments	$\searrow$ random edges, $\nearrow$ elliptic object, probably nucleus
aspect ratio of ellipse (b/a)	$\searrow$ random edges, $\nearrow$ elliptic object, probably nucleus
avg. level of texture inside ellipse	$\searrow$ probably nucleus, $\nearrow$ probably granulocyte or dust
variance of texture inside ellipse	$\searrow$ probably regular, $\nearrow$ probably pathological

Table 1. Extracted features.

3.3 Classification and visualization

Calculated features are provided to a trained classifier that assigns the tested path to one of two classes: abnormal nuclei (A_1) or other objects (A_2). Abnormal nuclei are: ASCUS, L–SIL, H–SIL. The other objects cover normal nuclei, nuclei–like objects and phantom objects. These two classes were chosen to distinguish objects that are under and out of interest. When the grouping process is completed, paths classified as A_1 represent final abnormal nuclei. If any abnormal nucleus is detected in an image, it is marked with a rectangle and the image is exposed for visual inspection. If more than one object is detected, all objects are marked. In the case of no detection, the image is rejected.

4 Experiments

Training and evaluation of classifiers were performed using Weka[1] environment [7]. The process of learning made no assumptions about the meaning of the particular features. The experimental set contained 2436 vectors of features classified by a medical expert (Tab. 2a). The little portion of A_1–class objects is caused by their rare occurrence in real images. Due to unbalanced training set the cost–sensitive learning was performed. Note, that only a few pattern recognition algorithms allow to minimize weighted cost, so in some cases instead of minimizing weighted cost, the training data instances were reweighted. The best results were obtained for cost factor $c = 32$. For most classifiers this value minimized the total and the weighted misclassification ratio. A comparison of several known classification methods with their standard WEKA parameters is given in Tab. 2b. Measures for evaluating classifiers were defined as follows: p_1,p_2 — misclassification ratio for A_1,A_2; p_t — total misclassification ratio; $p_w = (p_1 + p_2)/2$ — weighted misclassification ratio. Training and testing were executed using the cross–validation method. The experimental set was divided into four folds. In each pass three folds were used for training and one for testing. The correctness of classification ($1 - p_t$) is ranged between 90% and 95% depending on the applied classifier. The lowest total misclassification p_t is achieved for kNN methods, decision trees and rules (J48, RIP). To preserve low misclassification for class A_1, the measure p_w was chosen to evaluate cost sensitive classifiers. The kNN algorithm and linear regression algorithms are the leading ones due to their correctness and simplicity. Other methods (LMT, MP, NB) also keep the high correctness of classification. It means that the maximum possible class separation (90%–95%) is achieved for selected features. Further improvements of the algorithm should concentrate on searching for more distinctive features.

a)

Description	Quantity	Class
Pathological nuclei	41	A_1
Atypical nuclei	16	A_1
Normal nuclei	389	A_2
Luclei–like objects	178	A_2
Phantom objects	1812	A_2

b)

Algorithm	p_1 (%)	p_2 (%)	p_t (%)	p_w (%)
kNN(k=5)	7.02	3.36	3.45	5.19
Simple Linear Regression (SLR) [11]	3.51	7.73	7.64	5.62
Linear Regression (LR) [10]	5.26	6.73	6.69	5.99
Logistic model tree (LMT) [11]	5.26	7.10	7.06	6.18
Alternating Decision Tree (ADT) [13]	7.02	6.56	6.57	6.79
Multilayer Perceptron[2](MP) [12]	12.28	3.28	3.49	7.78
Naive Bayes Classifier (NB) [9]	7.02	8.62	8.58	7.82
kNN(k=10)	14.04	2.44	2.71	8.24
Naive Bayes Tree (NBT) [15]	10.53	6.18	6.28	8.35
Repeated Incr. Pruning (RIP) [16]	19.30	2.44	2.83	10.87
Pruned C4.5 Decision Tree (J48) [14]	26.32	2.02	2.59	14.17

Table 2. a) The expert classification for experimental set, b) cost sensitive classification (cost factor $c = 32$).

5 Summary

In this paper the algorithm was proposed to detect pathological cells in Ph cytological images. The algorithm could be useful for preliminary oncologic screening of cytological images because it is relatively fast and robust. Currently, the algorithm is being tested in a gynaecological clinic. An image processing stage of one 640x480 image takes from 4 up to 6 seconds on AMD ATHLON 1.8GHz, 512 MB RAM. A classification stage is immediate as the model and its parameters were fixed off–line. Examples of classification using trained SLR classifier are given in Fig. 3. Numerous edge segments resulting from various cell structures did not influence the correctness of classification significantly.

[1] Weka is the open source software for data pre-processing, classification, regression, clustering and visualization

[2] MP parameters: 1 hidden layer, 5 hidden nodes, 2 output nodes, Learning Rate=0.3, Momentum=0.2

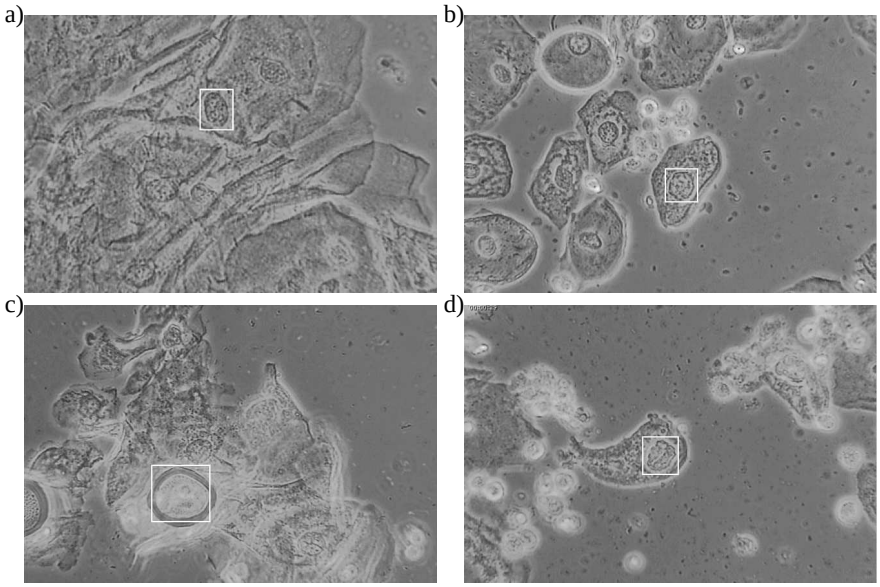


Figure 3. Exemplary 640x480 Ph images after processing and visualization.

References

1. G. Glab, K. Florczak, J. Jaronski, T. Licznerski, Gynaecological cyto–diagnosis in phase contrast microscopy (in Polish), Blackhorse Publ., Warszawa, 2001.
2. G. Glab, Evaluation of usefulness of joined phase-contrast microscopy and colposcopy examinations of uterine cervix in daily gynaecological practice. PhD. thesis (in Polish). Collegium Medicum, Jagiellonian University, Krakow, 1998.
3. M. Smereka, Nuclei recognition in phase contrast microscopy images, Proc. of the 3rd Int. Conf. on Computer Recognition Systems KOSYR'03 (2003) 35–40.
4. M. Smereka, Detection of elliptical shapes using contour grouping, Proc. of the 4th Int. Conf. on Computer Recognition Systems CORES'05 (2005) 443–450.
5. K. F. A. Ross, Phase contrast and interference microscopy for cell biologists, Edward Arnold Publ., London, 1967.
6. A. Fitzgibbon, M. Pilu, R. B. Fisher, Direct least square fitting of ellipses, IEEE Trans. on Pattern Analysis and Machine Intelligence 21 (5) (1999) 477–480.
7. I. H. Witten, E. Frank, Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
8. R. O. Duda, P. E. Hart, Use of the Hough Transform to detect lines and curves in pictures, Communications of the ACM 15 (1972) 11–15.
9. G. H. John, P. Langley, Estimating continuous distributions in Bayesian classifiers, Proc. of the 11th Conf. on Uncertainty in Artificial Intelligence (1995) 338–345.
10. S. Cessie, J. C. van Houwelingen, Ridge estimators in logistic regression, Applied Statistics 41 (1) (1992) 191–201.
11. N. Landwehr, M. Hall, E. Frank, Logistic model trees, ECML (2003) 241–252.
12. J. Hertz, A. Krogh, R. G. Palmer, Introduction to the teory of neural computation, Addison–Wesley Publ., 1991.
13. Y. Freund, L. Mason, The alternating decision tree learning algorithm, Proc. of the 16th Int. Conf. on Machine Learning (1999) 124–133.
14. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA., 1993.
15. R. Kohavi, Scaling up the accuracy of naive–Bayes classifiers: a decision tree hybrid, Proc. of the 2nd Int. Conf. on Knowledge Discovery and Data Mining (1996) 202–207.
16. W. W. Cohen, Fast effective rule induction, Proc. of the 12th Int. Conf. on Machine Learning (1995) 115–123.
17. *http://www.microscopyu.com/articles/phasecontrast/phasemicroscopy.html*

Imaging system design for detecting small changes in telomere length

A J McCollum*, C Restif and W F Clocksin

Department of Computing, Oxford Brookes University, Wheatley, Oxford OX33 1HX.

Abstract. This paper discusses some theoretical design considerations for an imaging system for telomere quantification in terms of Shannon channel capacity and reviews the practical issue of calibrating the measured signals. The eventual goal is to detect foetal cells in the maternal circulation based on telomere length. In designing a system to acquire and measure image features, there is a trade-off between bit-depth and spatial resolution. It is shown that the channel capacity of an image increases much more rapidly with the number of sensing elements than it does with the number of quantisation levels. Given the design choice between the two, it is more efficient to increase the spatial resolution of the sensor. The measurements are affected by variations in microscope illumination and set up, background noise, nonlinearity and drift in the instrumentation, hybridization efficiency, inconsistency of fluorophore standards between batches and in a given batch, and bleaching with exposure time. For rare cell detection, it may be more important to identify the rare events by a relative measure, than to quantify them absolutely. In this case, calibration internal to the nucleus, such as a centromeric probe, will be more reliable than the external methods.

1 Introduction

Invasive procedures for prenatal diagnosis involve risks to the foetus (1-2% rate of induced abortion) as well as giving maternal stress, discomfort and possible injury [1]. In the developed world, some 5% of pregnant women undergo such procedures. Non-invasive prenatal diagnosis is therefore an important objective. Although free foetal DNA is present in maternal plasma and serum, it can not be used to detect chromosomal conditions such as trisomy 21 (Down's Syndrome) or disorders where both parents are carriers of the same mutant allele. Such tests are the majority. A comprehensive diagnosis can only be carried out if foetal cells are available. It is useful to consider imaging system design for large-scale screening of cells.

Foetal cells are known to cross the placental barrier but are present in low concentrations (one foetal cell in $10^5 - 10^7$) in the maternal circulation during pregnancy [1]. Foetal cells obtained from maternal blood cannot be efficiently cultivated; therefore metaphase analysis cannot be performed; however, with interphase cells, some 70% of relevant chromosomal abnormalities could be covered with methods based on fluorescence in-situ hybridisation (FISH) and the polymerase chain reaction (PCR) [2]. Automated non-contact retrieval of single cells is already a practical proposition [3]. It may be possible in the future to isolate and culture living foetal cells individually. With appropriate image analysis software, various types of rare foetal cell might be isolated from maternal blood and collected with only a single preparation step.

The remaining challenges are then the design of robust and specific labels for the cells of interest, and the development of reliable automated image capture and processing systems that meet the diagnostic throughput rate required in a clinical environment. The reliability of rare cell detection depends on the label specificity and the signal to background ratio, while its speed depends on the label intensity. Even if the concentration can be enriched to one foetal cell in 10^4 maternal cells, say, automated cell identification will be needed if the diagnosis is to be performed in a reasonable time. This paper discusses some theoretical design considerations for an appropriate imaging system in terms of Shannon channel capacity and reviews the practical issue of calibrating the measured signals.

2 Telomeres as foetal cell markers

It has not yet been possible to use foetal cells for non-invasive diagnosis, partly due to their low incidence and partly because only certain types, in particular foetal erythroblasts, have been targeted for the purpose. If a bright, robust and specific label existed, that correctly identified foetal cells, then existing image processing techniques could be used without further development. No such marker has yet been found; however, it has been proposed that telomere length could be used as the basis of a "pan-foetal" cell marker to identify all foetal cells in the maternal circulation, regardless of type [1].

*Email: mccollum@brookes.ac.uk

2.1 Rates of telomere shortening

Human telomeres consist of approximately 10,000 base pairs (10 kbp) of tandem repeated sequences (TTAGGG), capping the ends of the chromosome arms. Telomeres prevent end-to-end fusion and are essential for chromosomal stability. Telomere length is known to decrease with age at an average rate of about 50 base pairs per year (bp/yr) from the age of 4 yr. At birth, the rate is orders of magnitude higher, at least 3 kbp/yr and 1 kbp/yr for granulocytes and lymphocytes respectively [4]. Foetal telomeres might therefore be expected to be conservatively 30% longer than those of the pregnant mother.

2.2 Labelling the telomere mass

Fluorescent marker dyes may be arranged to bind specifically to telomere sequences throughout the nucleus, enabling their locations to be imaged. An indication of telomere mass can be obtained from the dye emission, on the basis that the greater the telomere mass, the greater the quantity of binding probe. It is therefore necessary to quantify the fluorescent intensity emitted by the FISH spots. The image processing requirement, per nucleus, is to compute the total integrated fluorescent emission from the labeled material.

Telomeres are present at the two ends of the chromosome arms. In interphase there are 23 pairs of chromosomes, giving a total of 92 labelled fluorescent spots in the nucleus of a normal cell. FISH spot counting requires only that the nuclei of interest can be identified and that the spots are visible. For quantification of telomere mass, the strength of the probe signal must be measured. The emitted fluorescent light flux is expected to be greater from the probed foetal telomere mass, compared to that of the mother. In effect, the search for rare foetal cells in maternal blood is the search for outlying signal intensities.

Early work concentrated on metaphase nuclei, with their highly structured chromosomal arrangement [5], or on species that have much longer telomeres than humans, with a correspondingly strong probe fluorescence signal [6] [7]. In our case the fetal cells are in interphase where the DNA is distributed through the 3D volume of the nucleus. There is no visible chromosomal structure inside the nuclear boundary.

3 Imaging system design

Recent models of interphase chromatin suggest it is compacted corresponding to an estimated mass density of 110 to 150 base pairs per nm [8]. If this is the case, the 10 kbp typical of adult human telomeres might be expected to have physical length of about 0.1μm.

The diffraction limit of an ideal lens is proportional to NA/λ. With green light of $\lambda = 500$ nm, an oil immersion lens of NA = 1.25 gives a cutoff frequency of 5.0 cycles per micron, spreading the labelled telomere spot to a diameter of 0.2μm. In practice, defocus and scatter will increase the spread further. For telomere quantification, this is not necessarily a restriction because the requirement is to represent the energy radiated by the spot, rather than its shape.

It is informative to consider the design of a suitable imaging system in terms of channel capacity C [9]. For an Additive White Gaussian Noise (AWGN) channel,

$$C = W \log_2 \left(\frac{S}{N}\right) \tag{1}$$

where W is the system bandwidth, S is the output signal power (since the input signal is unknown), N is the output noise power and C is given in bits per second. In a given time interval, T, the maximum number of bits that can be delivered via the channel is $C_t = CT$. The quantity C_t is then given in bits (delivered over the time interval T).

In an imaging system, the channel can be partitioned into a series of three components corresponding to the optical, electronic (camera) and discrete (sampled) subsystems [10]. The optical channel capacity is limited by speckle noise, and is maximised in our case because the fluorescence is incoherent. The electronic channel capacity, C_e, is concerned with the conversion of photons to electrons:

$$C_e = XYB^2 \log_2 \left(\frac{S_e}{N_e}\right) \tag{2}$$

where X and Y are the dimensions of the sensor, S_e is the electronic output signal power and N_e is the electronic noise power. Here, the bandwidth, B, is the spatial sampling frequency of the sensor array, in cycles per unit distance. For example, a CCD sensor of dimensions $X = 8.58$ mm by $Y = 6.86$ mm, with 1280 by 1024 pixels respectively, has

$B = 1280/8.58 = 1024/6.86 = 149$ pixels/mm at the image plane (the sensor array). C_e is now given in bits per image. The electronic signal to noise ratio S_e/N_e is dependent on the photon flux and can be shown to be approximately the average number of photons impacting on each photosensor element [10].

The channel capacity of the sampled signal is proportional to the sensor area as well as the square of the spatial sampling rate, but is only logarithmically proportional to the electronic signal to noise ratio (SNR). Any attempt to improve the SNR by increasing the photon flux will increase the bleaching rate of the fluorescent dye, may increase the exposure time and will be weighted in logarithmic fashion. The implication is that it is much more efficient to increase either the sensor dimensions or the sampling frequency because these give a square-law improvement.

The number of elements in the sensor array is XYB^2. If the object of interest is an ideal point source, then, over the image, only one element is irradiated and the channel capacity is independent of either the sensor area or the sampling frequency.

$$C_{ep} = \log_2 (\frac{S_e}{N_e}). \tag{3}$$

Although this suggests that a point (or area) sensor could measure the intensity of a single FISH spot (or a group of such spots), in practice a 2D image is required to segment the stained nuclei and FISH spots before the measurement is made [11] [12]. Nyquist theory gives the minimum sampling rate that will completely represent the original image, given its diffraction limit. If W is the diffraction limited bandwidth of the optical system, then the spatial sampling frequency of the sensor is $B = 2W$. Under these conditions the channel capacity of the sensor is

$$C_e = 4XYW^2 \log_2 (\frac{S_e}{N_e}). \tag{4}$$

The required number of elements in the two dimensions of the sensor array are $M = 2XW$, and $N = 2YW$. If the image is quantised into g grey levels, then the channel capacity of the discrete channel is:

$$C_d = MN \log_2 (g). \tag{5}$$

Again, the channel capacity increases more efficiently with the number of pixels than with the number of grey levels. This is an important observation because it indicates that the payoff in increasing the spatial resolution is significantly greater than that of increasing the number of quantisation levels by the same amount.

4 Image calibration

Probe signals are affected by variations in microscope illumination and set up, background noise, nonlinearity and drift in the instrumentation, hybridization efficiency, inconsistency of fluorophore standards between batches and in a given batch, and bleaching with exposure time [13]. In practice, variations in label intensity are observed inside a given nucleus, across the field of view (typically 0.14 mm by 0.1 mm at 63x magnification) and across the entire microscope slide. Calibration of the measurements is therefore required to compensate for these effects.

Table 1 summarizes five calibration techniques. They are classified here as external or internal methods. External methods are designed to compensate for variations in the instrumentation but are not concerned with effects in the nucleus that might locally affect the telomere probe signals. Such methods are convenient to use and give a traceable absolute reference standard, rather than a relative measure. Internal methods, as well as compensating for variations in the instrumentation, account for local effects in the nucleus such as changes in hybridization efficiency. Internal methods currently in use give only a relative measure of probe signal strength.

4.1 Background Correction

Flat field or background correction is provided as standard by commercial microscopy imaging systems. Calibration involves placing one or more featureless reference slides in the field of view and capturing the corresponding reference images. The reference is then used to correct further images taken during the experiment. As implemented currently, background correction is a relative rather than absolute normalization method. If fluorescent standards are used as references, then an absolute intensity calibration is possible. Background correction is, of course, external to the cell biology and will not compensate for effects at that level. It operates at a scale that is suitable for correcting images for display purposes. In our experience, at the magnifications used in telomere probe imaging, there are blemishes in the calibration image that introduce additional spatial noise.

Method	Advantages	Disadvantages
Background Correction (External)	No additional cost	Relative measure Decoupled from nucleus
Fluorescent Beads (External)	Absolute standard	Decoupled from nucleus
Plasmid references (External)	Absolute standard Gives mass in base pairs	Cumbersome in practice Decoupled from nucleus
Nucleic Staining (Internal)	No additional cost Directly coupled to nucleus	Relative measure Questionable reliability
Chromosomal Staining (Internal)	Directly coupled to nucleus interior	Relative measure High materials cost Difficult procedure

Table 1. Summary of Calibration Methods.

4.2 Fluorescent Beads

Kits of fluorescence standards, in the form of microscopic beads with known emissivity and size, are available from commercial suppliers. These can be used as a quantifiable reference. The beads are contained in an appropriate medium and spread on a microscope slide. Because the beads are distributed in 3D space (as are the telomere probes), some will be out of focus to an extent that prevents reliable quantitative intensity measurement. A stack of images is therefore acquired at different heights, either by vertically moving the microscope stage or by changing the focus of the microscope objective lens. A composite image is then generated by integrating along the z-axis. In practice, intensity measurement, unlike dimensional measurement, tends to be insensitive to small changes in focus. Calibration images are acquired before (and possibly during and after) the experiment. Statistics acquired from the calibration images are subsequently used to normalize the experimental results [6].

4.3 Plasmid References

Plasmids are loops of DNA that are maintained in the cell cytoplasm and can be engineered to have specific telomere repeat sequences. Plasmids are stable in that the telomere sequence does not change during cell division. Hybridized plasmids can therefore be used to correlate telomere fluorescence intensity with respect to length [6]. Although this method gives an absolute measure of telomere mass in terms of base pairs, it takes many hours to prepare slides and to generate results and is therefore impractical to use in routine tests. It can, however, produce a characteristic that can be subsequently corrected using fluorescent bead standards, which are much more convenient to use. This combination of methods gives a routine experimental method for quantifying telomere sequences in base pairs. The technique is still external to the nucleus in which the measurement is made, and cannot therefore compensate for local effects.

4.4 Nucleic Staining

For measurements in interphase, where the chromosome arrangement is unstructured, a counterstain such as DAPI (4',6-diamidino-2-phenylindole) is used to mark the nucleus. To reduce crosstalk between the telomere probe signal and that of the counterstain, their characteristic spectra are chosen to be well spaced and without overlap, and are separated by optical filters. Some workers have used the counterstain signal as a calibration reference [14] [15]. Since DAPI staining is part of the routine procedure, no additional cost is incurred. As the staining is internal to the nucleus it might be expected to compensate for local effects at the chromosomal scale (as well as for global instrumentational variations). Our experience is that the relationship is not reliable in practice.

4.5 Chromosomal Staining

A ratio of a probe signal and a local nucleic reference will tend to cancel the common-mode effects inside the nucleus. Such an internal reference can be produced by probing a known sequence, for example the centromeres of specific chromosomes [16]. Markers localised in the same chromosome can be assumed to have similar hybridisation efficiencies [17]; however, their bleaching curves should also be similar or at least available to be included in the calibration procedure. Chromosomal staining can compensate for all of the sources of uncertainty mentioned earlier; however, it is a relative measure, high cost, and in our experience a difficult procedure to perform. To create a PNA FISH probe, a sequence of about 200 minisatellites must be assembled. Due to the constraints of PNA synthesis, each minisatellite cannot be greater than 20-25 base pairs.

5 Conclusions

In designing a system to measure telomere spot intensity, there is a trade-off between bit-depth and spatial resolution. The channel capacity of an image increases much more rapidly with the density of sensing elements than it does with the number of quantisation levels. Given the design choice between the two, it is more efficient to increase the spatial resolution of the sensor.

An ideal calibration method for quantitative FISH telomere signal measurement would compensate for drift and other variations in the instrumentation, as well as for local effects at the nucleus, would be traceable to a reference standard, and would give an absolute value in base pairs. It can be seen from Table 1 that none of the methods combine all of the ideal characteristics. Internal methods have the potential to account for hybridisation efficiency and other effects local to the nucleus (as well as compensating for variations due to the instrumentation) but give a relative rather than absolute measure. External calibration methods, in principle, are traceable to an absolute standard but are intrinsically incapable of compensating for internal nucleic variation.

For rare cell detection in a clinical environment, such as the detection of one foetal cell in 10^4 maternal cells, it may be more important to identify the rare events than to quantify them. In this case, internal calibration such as centromeric probing, will be more reliable than external methods.

Acknowledgements

Acknowledgements are due to Professor M Hulten and Dr Y Ariosa of the University of Warwick. This work was supported by a European Commission grant project SAFE 503243.

References

1. S. Hahn & W. Holzgreve. *Fetal cells and Fetal DNA in maternal blood*. Karger, Basel, 2001.
2. S. Hahn, R. Sant & W. Holzgreve. "Fetal cells in maternal blood: current and future perspectives." *Molecular Human Reproduction* **4**, pp. 515–521, 1998.
3. R. Burgemeister. "New aspects of laser microdissection in research and routine." *J Histochem & Cytochem* **53**, pp. 409–412, 2005.
4. N. Rufer, T. Brummendorf, S. Kolvraa et al. "Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood." *J. Exp. Med* **190**, pp. 157–167, 1999.
5. U. Martens, J. Zijlmans, S. Poon et al. "Short telomeres on human chromosome 17p." *Nat. Genet.* **18**, pp. 76–80, 1998.
6. S. Poon, U. Martens, R. Ward et al. "Telomere length measurements using digital fluorescence microscopy." *Cytometry* **36**, pp. 267–278, 1999.
7. M. Hande, E. Samper, P. Lansdorp et al. "Telomere length dynamics and chromosomal instability in cells derived from telomerase null mice." *J Cell Biol* **144**, pp. 589–601, 1999.
8. K. Bystricky, P. Heun, L. Gehlen et al. "Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques." *PNAS* **101**, pp. 16495–16500, 2004.
9. C. Shannon & W. Weaver. *The mathematical theory of communication*. University of Illinois Press, Illinois, 1949.
10. C. Wagner & G. Hausler. "Information theoretical optimization for optical range sensors." *Applied Optics* **42**, pp. 5418–5426, 2003.
11. C. Restif & W. Clocksin. "Comparison of segmentation methods for cytometric assay." In *Medical Image Understanding and Analysis*, pp. 153–156. London, UK, September 2004.
12. C. Restif. "Towards safer, faster prenatal genetic tests: Novel unsupervised, automatic and robust methods of segmentation of nuclei and probes." In A. Leonardis, H. Bischof & A. Prinz (editors), *European Conference on Computer Vision*, volume LNCS 3954, pp. 437–450. Springler-Verlag, Graz, Austria, May 2006.
13. S. Poon & P. Lansdorp. "Quantitative fluorescence in situ hybridization (Q-FISH)." In *Current Protocols in Cell Biology*, pp. 18.4.1–18.4.21. John Wiley & Sons, Inc, 2001.
14. E. D. Pauw, N. Verwoerd, N. Duinkerken et al. "Assessment of telomere length in hematopoietic interphase cells using in situ hybridization and digital fluorescence microscopy." *Cytometry* **32**, pp. 163–169, 1998.
15. R. Narath, T. Lorch, K. Greulich-Bode et al. "Automatic telomere length measurements in interphase nuclei by IQ-FISH." *Cytometry Part A* **68**, pp. 113–120, 2005.
16. R. Narath, T. Lorch, M. Rudas et al. "Automated quantification of gene amplification in clinical samples by IQ-FISH." *Cytometry Part B* **57**, pp. 15–22, 2004.
17. P. Perner, S. Bruderlein, C. Hasel et al. "Quantifying telomere lengths of human individual chromosome arms by centromere-calibrated fluorescence in situ hybridization and digital imaging." *Americ J Pathol* **163**, pp. 1751–1756, 2003.

Deformable Image Registration using Spring Mass System

Jian-Kun Shen[a], Bogdan J. Matuszewski[a], Lik-Kwan Shark[a], Christopher J. Moore[b]

[a]Applied Digital Signal & Image Processing Research Centre, University of Central Lancashire, Preston, UK
[b]North Western Medical Physics, Christie Hospital NHS Trust, Manchester, UK

Abstract. The paper describes a novel multi-resolution registration method. The method is fast, robust and offers high registration accuracy. The algorithm models deformations using an elastic spring mass system, which contains sparse masses interconnected by springs. The proposed method uses data intensity values to guide deformation with local constraints imposed by interaction of inter-connecting springs. Moreover, by using such system prior information about the data can be easily taken into account to improve the registration accuracy. The performance of the method is tested using simulated as well as real dynamic magnetic resonance image data.

1. Introduction

Image registration has become central in medical image analysis [1,2]. Much research has been done in this area in the last twenty years. Whilst global rigid/affine registration has matured, deformable registration is still under intense development. The registration methods can be broadly divided into two main categories, feature-based and intensity-based methods. By matching the corresponding image features extracted from images, the feature-based methods calculate deformation of the whole image using one of "smooth" interpolation methods. The intensity-based methods operate directly on image intensity values. One of the most popular methods is to calculate the transformation using a set of equally spaced sparse control points, which are not linked to any specific image features, and finding the extreme of the similarity measure defined in the neighbourhood of the control points. Subsequently, the image deformations are calculated from displacement of sparse control points using one of "smooth" interpolation methods. As the number of control points might be significant additional regularisation measures are necessary to avoid excessive variation of the deformation field [3]. Since the intensity-based methods estimate the transformation based on the entire image contents, these methods are more computationally demanding. Whilst feature-based methods can greatly reduce computational complexity, they also have number of disadvantages including: (i) the feature extraction methods produce inherent erroneous features due to imperfection and noise in the image acquired; (ii) the registration errors can increase due to uneven feature distribution. This paper describes a hybrid method. In the proposed method, the deformation is modelled using an elastic spring mass system, containing a number of sparse masses and springs interconnecting them. At the initial stage it is assumed that the system of interconnected masses is in equilibrium. One of the advantages of using such a system is that the system can explicitly represent features present in the image in order to improve the registration accuracy. By moving masses representing the features in one image to the corresponding location in the other image, the remaining masses will change position until all the forces are in equilibrium. In the following stage, the external forces calculated from the image contents are introduced into the system to locally guide movable masses to gradually improve registration accuracy under the constraints defined by the internal forces. The applied multi-resolution approach reduces the computational time and improves the registration accuracy. The described method can be interpreted as a version of the Free Form Deformation (FFD) registration method.

The remainder of the paper is organised as follows: Sections 2, 3 and 4 briefly describe three main steps in the algorithm. They are spring mass system generation, motion modelling, and image warping; Section 5 describes a multi-resolution registration scheme; in Section 6 the registration results are presented and the conclusions are drawn in Section 7.

2. Spring Mass System Generation

The spatial topology of the spring mass system is defined using a mesh generation algorithm with the vertices of the mesh interpreted as the masses and the mesh edges as the springs connecting these masses. To achieve high registration accuracy, the mesh generation algorithm should provide: *(i)* Well shaped triangles (all triangles should have three inner-angles close to 60°) to reduce sensitivity to the mesh orientation; *(ii)* Variable mesh density that is controllable by a user; *(iii)* Vertices coinciding with all feature landmarks; *(iv)* Good representation of the feature curves (such as anatomical structures) with the vertices placed on the curve and mesh edges lying along the curve with controllable approximation errors. Any mesh generation algorithms can be used as long as they could fulfil these requirements. In this paper, the bubble mesh [4] has been used to generate the mesh vertices. More details about application of the bubble mesh method for generation of the spring-mass system can be found in [5,6]

3. Motion Modelling

The motion of the system is described here by the Newtonian mechanics and is expressed by a set of *2-nd* order Ordinary Differential Equations. The *4-th* order Runge-Kutta method has been used to yield the numerical solution.

$$m\frac{d^2\mathbf{p}_{(i)}}{dt^2}+c\frac{d\mathbf{p}_{(i)}}{dt}=a(t)\cdot\mathbf{f}_{\mathbf{int}(i)}+b\cdot\mathbf{f}_{\mathbf{ext}(i)} \qquad \bigwedge_{i\in\mathcal{N}} \tag{1}$$

where:

- $\mathcal{N}$ - is a set of indexes for which the corresponding masses in the spring mass system are movable.
- $\mathbf{p}_{(i)}=[x_i \;\; y_i]^T$ - is the position of the *i-th* movable mass.
- $\mathbf{f}_{\mathbf{int}(i)}$ - is the internal force caused by the springs connected to the *i-th* mass.
- $\mathbf{f}_{\mathbf{ext}(i)}$ - is the external force acting on the *i-th* mass.
- $a(t)$- is iteration dependant scalar which weights the contribution of the internal and external forces. Initially the value of scalar $a(t)$ is set to one. Subsequently, after the external force is activated, the value of this parameter is decreasing gradually from one to the threshold value of a_{min}.
- b - is a binary value which controls whether the external forces are acting on the movable masses. In the initial stage, $a(t)$ is set to one and b is set to zero to disable the external forces. Subsequently, b is set to one to enable the external forces and $a(t)$ starts to decrease. Consequently, the external forces calculated from the image contents can guide the deformation.

3.1 Internal Forces

The role of internal forces in the system is to make sure the calculated deformation is "smooth". In the traditional spring force formula [5], the repulsive force can cause unexpected results when the deformable images contain large local movement. Therefore, the repulsive force needs to be removed. To keep the system stable in the initial stage (meaning that $\mathbf{f}_{\mathrm{int}(i)}=0$), the extra attractive force is added locally along each spring direction. If a set of indexes for which the corresponding masses are connected to the current *i-th* mass is defined as $\mathcal{M}(i)$, the internal force [6] can be written in general as:

$$\mathbf{f}_{\mathbf{int}(i)}=\sum_{j\in\mathcal{M}(i)}k_j\cdot\frac{\mathbf{p}_j-\mathbf{p}_i}{\|\mathbf{p}_j-\mathbf{p}_i\|}\cdot\left(\|\mathbf{p}_j-\mathbf{p}_i\|+\rho_{ji}\right) \qquad \bigwedge_{j\in\mathcal{M}(i)} \tag{2}$$

where:

- k_j - is the spring constant, which locally controls how easily the system can be deformed.
- $k_j\rho_{ji}$ - is the "added" force acting along the *ji* spring direction. It is important to notice that in general $\rho_{ji}\neq\rho_{ij}$.

The scalars ρ_{ji}, where $j\in\mathcal{M}(i)$, need to be calculated. However, there are infinite number of possible selections for ρ_{ji}. In the method described in this paper the unique solution is defined as the solution a constraint optimisation problem where the cost function is given by:

$$\arg\min_{\rho_{ji}}\left(\sum_{j\in\mathcal{M}(i)}\left(k_j\cdot\rho_{ji}\right)^2\right) \tag{3}$$

subject to equality constraint in (4) and inequality constraints in (5)

$$\mathbf{f}_{\mathbf{int}(i)}=0 \tag{4}$$

$$\rho_{ji}\geq 0, \qquad \textit{where } j = 1 \ldots N \tag{5}$$

This problem can be solved using Lagrange Multiplier method [6].

3.2 External Forces

The external force $\mathbf{f}_{\mathbf{ext}(i)}$ is a force indicating how much the *i-th* mass should be moved towards the "good" matching direction. The role of external forces is to incorporate information of image contents into the system to achieve iteratively higher and higher registration accuracy. In this case, a small size image I_i^w , with the centre located at the i-*th* mass, is extracted from the warped source image to estimate the good local matching direction in the target image I_i^t. The external force is defined as:

$$\mathbf{f}_{\mathbf{ext}(i)} = \beta \cdot grad\left(f\left(d\left(I_i^w, I_i^t \right) \right) \right) \tag{6}$$

where: $d(\cdot,\cdot)$ is a similarity measure between two images and β is a measure of the reliability of the gradient.

The proposed algorithm enables the use of different similarity measures such as sum of square differences or mutual information. For the results presented in this paper, the normalized cross correlation function in equation (7) has been used. This similarity measure performs well as long as a relation between source and target image intensities can be approximated by the affine transformation.

$$f\left(d(\cdot,\cdot)\right) = \begin{cases} A \times \left(d(\cdot,\cdot)\right)^n & d(\cdot,\cdot) \geq 0 \\ 0 & d(\cdot,\cdot) < 0 \end{cases}, \quad \text{where } A = 50 \text{ and } n = 2 \tag{7}$$

4. Image Warping

After the solution of equations in (1) has been found the source image needs to be warped to match the target image. This process involves coordinate transformation to find corresponding locations of all the pixels from the warped source image in the source image (inverse mapping is used) and subsequent computation of the pixel values in these new locations. While the standard bilinear interpolation is used for the latter, one of interpolation methods for scattered data needs to be used for the former. The image warping needs to be performed during each iteration of the algorithm. Therefore, the piecewise affine interpolation based on Constrained Delaunay Triangulation [5] is used.

5. Multi-resolution Registration Scheme

The multi-resolution approach [7] is an often used technique to register the images with larger deformation. In the work presented in the paper, a Gaussian pyramid is built for the source and target images first. Subsequently, the registration scheme starts at the coarsest level and progress to the finest level. The multi-resolution registration scheme is summarised in figure 1. In the work, *d_rati*o is set to 0.80 and the threshold a_{min} is equal to 0.3;

Start with the lowest resolution data in the pyramid decomposition;
Build the spring mass system with good representation of features (such as landmarks, lines, curves) in the source image;
Move the masses representing the image features to their corresponding locations in the target image;
Model the deformation using the spring mass system without external forces (*b=0*);
repeat
    a=1;
    while ($a > a_{min}$)
        Warp the source image;
        Model the deformation using the spring mass system with external forces (*b=1*);
        a = a · d_ratio;
    end
    Increase the data resolution;
    Build the spring mass system with good representation of features in the target image;
    Calculate the masses location in the source image using the transformation estimated in the previous pyramidal level;
until (not the top of the pyramid decomposition)

Figure 1. Multi-resolution Registration Scheme using Spring Mass System

It is typical for the multi-resolution registration to gradually improve on the displacement field estimate when progressing through the resolution levels. An example of such behaviour is shown in the next section. Typically a single resolution technique is prone to fail when large displacements are present in the data.

6. Results

The proposed method is first tested using simulated data. The 512-by-512 source image is warped using a sequence of randomised transformations to generate a set of simulated target images (see figure 2). Subsequently, the described registration algorithm is used to recover the known displacement field. To generate these simulated target images, the bony structures and the rectum have been segmented in the source image and each of them have been transformed rigidly with randomly selected translation vector (with maximum translation of ±20 pixels) and rotation angle (with maximum of ±15°). The simulated images are generated using method described in [8]. Figure 3 shows typical results obtained using the multi-resolution registration technique described in the previous section.

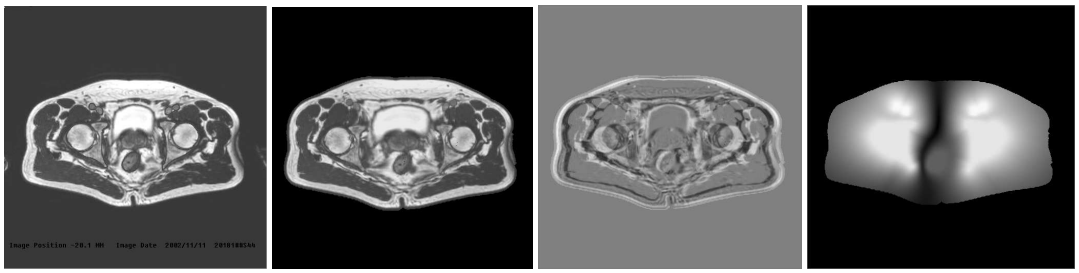


Figure 2. From left to right: source image with the segmented structures used for the simulation of the target image; one of simulated target images; difference between source and target image; displacement field between source and simulated target image.

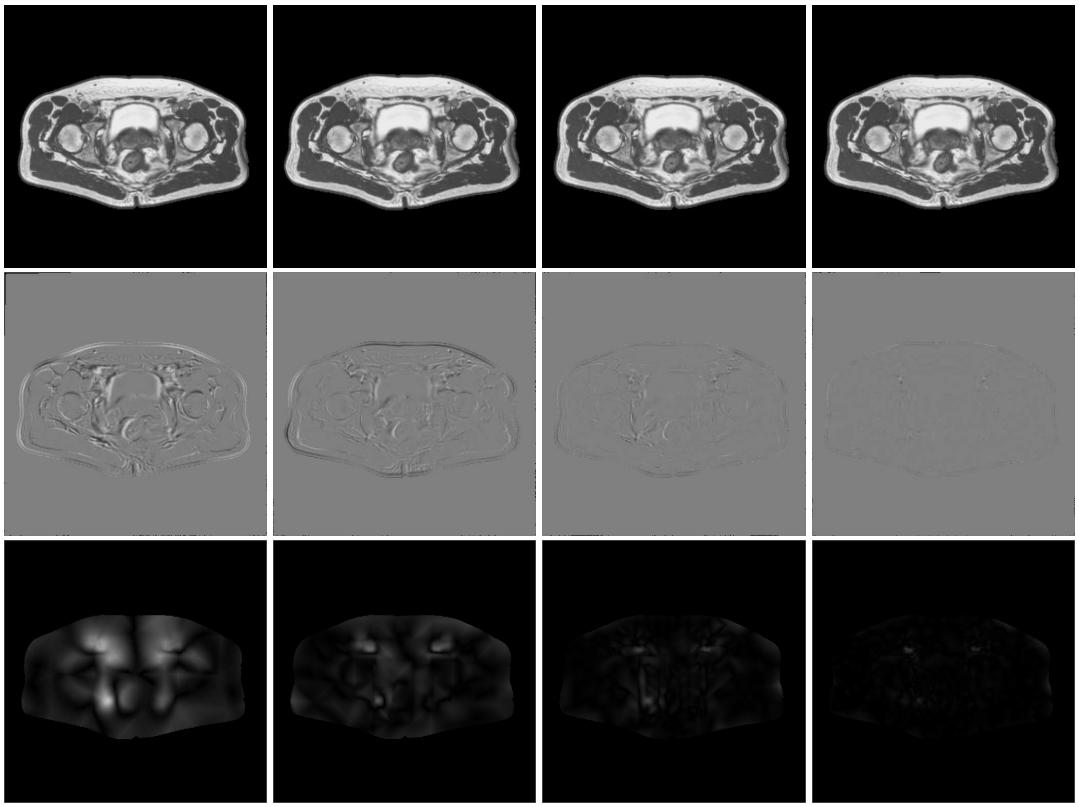


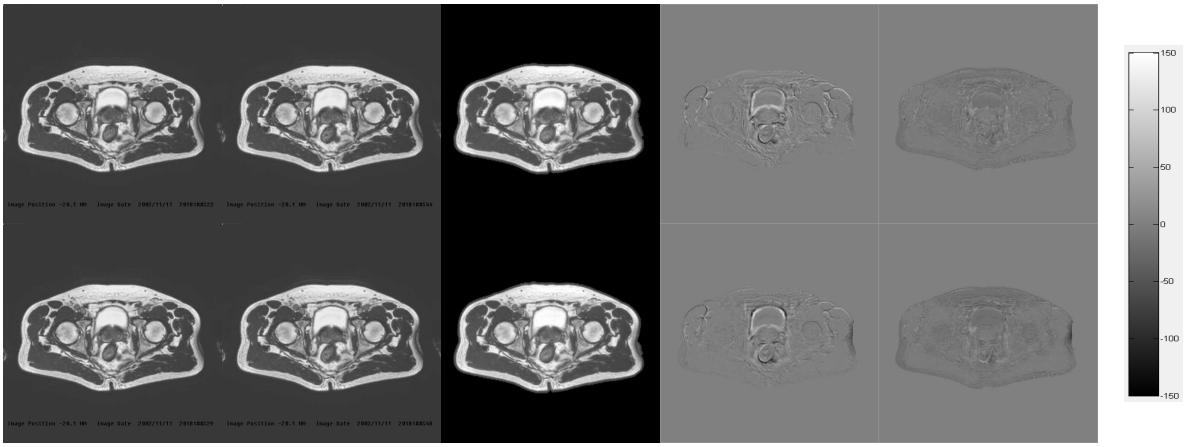
Figure 3. From top to bottom: registered source image, difference between registered and simulated target images, displacement field between registered and simulated target image. The results have been obtained at different resolution levels, as described in section 5 with the columns from left to right showing results obtained with one-eighth, one-fourth, half, and full resolution.

A simple statistics of the displacement field calculated for each resolution level are shown in Table 1. The column with bold letters in table 1 represent statistics calculated for the original simulated displacement field. The results shown in figure 3 and table 1 are typical for the all the simulated target images as described at the beginning of this section.

	Before	1/8 resolution	1/4 resolution	1/2 resolution	Full resolution
Max	**22.3607**	15.2603	8.1379	5.3598	3.9656
Mean	**12.8540**	2.8896	1.4601	0.6671	0.2832
std	**5.3509**	2.2378	1.0104	0.5703	0.3006

Table 1. Validation results using simulated images

The proposed method has been also applied for a sequence of real dynamic magnetic resonance images. The source and target images are randomly selected from the sequence. Two examples of the source, target and registered images are respectively shown in figure 3 (a), (b) and (c). Figure 3 (d) and (e) show differences between the source and target and between the registered and target images respectively.



(a) (b) (c) (d) (e)

Figure 3. Validation results using real dynamic magnetic resonance image data

7. Conclusion

The paper presents a new registration method using an elastic spring mass system, which enables to reflect prior information about image including spatial topology of the image features and local elastic properties. By moving the masses representing the features in one of the images to the corresponding location in the other image, an initial deformation is modelled via the internal forces caused by springs. Subsequently, the external forces are calculated from image contents and added into the equations. This guides the system to achieve high registration accuracy under the constraints of the internal forces. With the multi-resolution registration scheme, it improves registration quality for the large deformation. Although the registration results has been shown only using dMRI data in the same modality, the algorithm generally can be used to register multi-modality and multi-dimension data.

References

1. W.R. Crum, T. Hartkens and D.L.G. Hill "Non-rigid Image Registration: Theory and Practice", *The British Journal of Radiology*, Vol. 77, pp. 140-143, 2004.
2. D.J. Hawkes, *et. al*. "Tissue Deformation and Shape Models in Image-guided Interventions: A Discussion Paper", *Medical Image Analysis*, Vol. 9, pp. 163-175, 2005.
3. D. Rueckert, *et. al*. "Nonrigid Registration using Free Form Deformations: Application to Breast MR Images", *IEEE Transactions on Medical Imaging*, Vol. 18, No. 8, pp. 712-721, 1999.
4. K. Shimada and D.C. Gossard "Bubble Mesh: Automated Triangular Meshing of Non-manifold Geometry by Sphere Packing", *ACM 3rd Symposium on Solid Modelling and Applications*, pp. 409-419, 1995.
5. B.J. Matuszewski, J.-K. Shen and L.-K. Shark "Elastic Image Matching with Embedded Rigid Structures using Spring Mass System", *IEEE International Conference on Image Processing (ICIP)*, Vol. 10, pp. 937-940, 2003.
6. J.-K. Shen, B. J. Matuszewski and L.-K. Shark *"Deformable Image Registration"*, *IEEE International Conference on Image Processing (ICIP)*, pp 1112-15, 2005.
7. R. Bajcsy and S. Kovacic "Multiresolution Elastic Matching", *Computer Vision, Graphics and Image Processing*, Vol. 46, pp. 1-21, 1989.
8. J.A. Little, D.L.G. Hill, and D.J. Hawkes, "Deformations Incorporating Rigid Structures", Computer Vision and Image Understanding, Vol.66, No. 2, pp. 223-232, 1997.

Malaria Parasite Detection in Peripheral Blood Images

F. Boray Tek [a], Andrew G. Dempster[b] and Izzet Kale [a,c*]

[a]University of Westminster, Applied DSP & VLSI Research Group, Dept. E. Sys, London, UK
[b]University of New South Wales, School of Surveying and Spatial Information Systems, Sydney, Australia
[c]Eastern Mediterranean University, Applied DSP & VLSI Research Centre, Dept. Eng, N. Cyprus

Abstract. This paper investigates the possibility of computerised diagnosis of malaria and describes a method to detect malaria parasites (*Plasmodium spp*) in images acquired from Giemsa-stained peripheral blood samples under conventional light microscopes. Prior to processing, the images are transformed to match a reference image colour characteristics. The parasite detector utilises a Bayesian pixel classifier to mark stained pixels. The class conditional probability density functions of the stained and the non-stained classes are estimated using the non-parametric histogram method. The stained pixels are further processed to extract features (histogram, Hu moments, relative shape measurements, colour auto-correlogram) for a parasite/non-parasite classifier. A distance weighted K-nearest neighbour classifier is trained with the extracted features and a detailed performance comparison is presented. We show that our method achieves 74% sensitivity, 98% specificity, 88% positive prediction value, and 95% negative prediction value for the parasite detection.

1 Introduction

Malaria is a serious disease caused by a blood parasite named *Plasmodium spp* (henceforth parasite). The World Health Organization estimates 300-500 million malaria cases and more than 1 million deaths per year [9]. The definitive diagnosis of malaria infection is done by searching for the parasites in blood slides (films) through a microscope. However, this is a routine and time consuming task. Besides a recent study in the field shows the agreement rates among the clinical experts for the diagnosis are surprisingly low [8]. Hence, it is necessary to produce a common standard tool to perform diagnosis with same ground criteria uniformly everywhere.

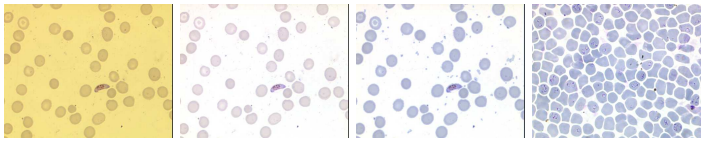
Automatic parasite detection has been addressed in [10], [11], [12] with thresholdings based on colour histograms. In [10], the image was pre-processed to remove the illumination bias which was estimated as the difference between the histogram peak positions of the colour channels (i.e. peak of blue - peak red/green). The markers for the parasites were extracted using a threshold at the halfway point of the cumulative histogram. However, in a diagnosis scenario, for every sample, such an approach would fail by marking all white blood cells (WBC), platelets and artefacts as parasites.

In this study we have proposed a solution for the parasite detection problem with two consecutive classifications: stained/non-stained pixel classification and parasite/non-parasite classification. Provided enough samples are present to estimate class conditional density functions, the Bayesian decision rule can be used as a powerful pixel classification method [2], [6], [7]. We have employed the Bayesian decision rule to determine if a pixel has stained or non-stained colour. The pixels that are classified as stained are further processed to form labelled connected components which are parasite candidates. Then these candidates are classified in a distance weighted K-nearest neighbour classifier to determine if they are parasites. For this classification, we have selected four different candidate features which are all rotation and scale invariant. We have evaluated and compared the performance of the individual selected features as well as their concatenated forms.

2 Colour Normalisation

Since the classifiers utilise colour information, it is essential to apply colour normalisation to the images to decrease the effect of different light sources or sensor characteristics. To perform efficient colour normalisation, blood slide images can be seen as a combination of two basic components (plasma and the rest) which can be separated by foreground-background segmentation. Hence, we have used a simple colour normalisation method which exploits this property of the images [15]. First, the input image was separated into foreground and background regions. Then, for each colour channel, the global illumination effects were reduced by scaling with the background channel averages. The normalisation was completed by transforming the foreground pixels according to a reference set. The procedure is demonstrated with an example image in Figure 1. The details of the procedure can be found in [15].

*e-mail: boraytek@yahoo.co.uk, a.dempster@unsw.edu.au, kalei@westminster.ac.uk.



(a) (b) (c) (d)

Figure 1. Colour normalisation: (a) An input image before normalisation (b) after normalisation with its background channel average values, (c) (b) normalised with reference foreground channel averages, (d) an example from the reference set.

3 Stained/Non-Stained Pixel Classification

Determining the stained pixels can be viewed as a two class classification problem. Suppose that the two classes are w_s for the stained and w_{ns} for the non-stained. A Bayesian classifier for minimum cost with a feature x in the form of **rgb** colour vector can be formulated as follows [2]:

$$\mathbf{rgb} \in w_s \qquad if \qquad \frac{p(\mathbf{rgb}|w_s)}{p(\mathbf{rgb}|w_{ns})} \geq \theta, \tag{1}$$

where $p(\mathbf{rgb}|w_i)$ ($i \in \{s, ns\}$) denotes the class conditional probability density function. θ can also be understood as a threshold for the likelihood ratio. It represents implicitly the application dependent costs for the decision and especially the *a priori* probabilities of classes $P(w_s)$, $P(w_{ns})$ when they are not easily determinable [2].

There are parametric, non-parametric and semi-parametric methods to estimate the $p(\mathbf{rgb}|w_i)$ [16]. We have used a non-parametric method based on histograms [6]: the probability density functions can be obtained from the histograms of the labelled samples:

$$p(\mathbf{rgb}|w_s) = \frac{H_s(r,g,b)}{N_s} \qquad p(\mathbf{rgb}|w_{ns}) = \frac{H_{ns}(r,g,b)}{N_{ns}} \tag{2}$$

H_i and N_i denote the histogram which counts the number of the occurrences of (r, g, b) triples and the total number of the samples for class w_i ($i \in \{s, ns\}$), respectively.

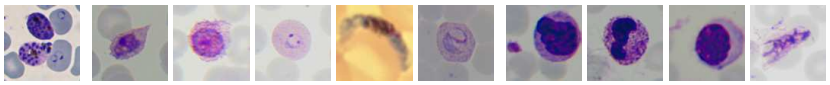
4 Parasite/Non-Parasite Classification

In addition to the parasites the stained pixels can belong to other components such as WBCs, platelets or artefacts. To finalise the detection further classification is required. However, the stained pixels have to be represented as stained objects (connected sets) to extract features for the classifier. Furthermore, the regions that include the objects have to be found. This is achieved with the tophat extraction and infinite reconstruction which are based on mathematical morphology [13]. Additionally, calculating the area pattern spectrum ($G_\Lambda(X)$) function on the negative of the grey level image provides an estimate for the average RBC area (A_μ) [10], [14]. A_μ is found by calculating the peak index of the $G_\Lambda(X)$: the differential plot of the reduced volume is calculated by successive area openings by an increasing area parameter. Let $\gamma^a_{\lambda_i}(X)$ denotes area opening of an image X with area threshold λ_i, then the area granulometry can be calculated as in (3):

$$G_\Lambda(X) = \sum \gamma^a_{\lambda_i}(X) - \sum \gamma^a_{\lambda_{i-1}}(X) \qquad \Lambda = \{\lambda_1, \lambda_2, ..., \lambda_n\} \tag{3}$$

4.1 Feature Extraction

Even for humans, parasite/non-parasite classification is not an easy task without special training. The parasite is a non-rigid object that can have a large variation in the observed morphology (Figure 2). The colour information is valuable but may not be adequate to distinguish the parasites from other stained objects. Raw images can not be used directly as features due to high variations in morphology which are coupled with arbitrary rotations and scales. A feature to be used in this classification must provide rotation, scale invariance and must be capable of capturing the morphological characteristics. We have chosen 4 different candidate features: (colour histogram, Hu moments, colour auto correlogram (henceforth correlogram), and a relative shape measurements vector) to investigate their individual performances and then search for a higher combined feature performance. The histogram is a widely used descriptor which is simple to compute and gives adequate information about the colour distribution. Hu's moment invariants are derived from algebraic combinations of the first 3 orders of normalised central moments. They are also rotation and



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 2. Examples of stained objects: (a)-(f) Plasmodium, (g)-(i) WBCs (g) includes a platelet (j) artefact.

scale invariant while providing spatial information [4]. The third feature, what we call the relative shape measurements vector, is formed of simple measurements to represent the object shape. The colour auto correlogram can be seen as an extended histogram: carries spatial information in addition to colour distribution. It has been proposed in [5] for image indexing and sub-region localisation and was used in [1] for object recognition.

For every stained object S, four different features are extracted. These features are colour histogram ($\mathbf{H}$), Hu moment invariants ($\mathbf{M}$), relative shape measures ($\mathbf{R}$), colour auto correlogram ($\mathbf{C}$).
Colour histogram $\mathbf{H}$ is the count of the occurrences of the colour c_i in image I:

$$H(c_i) \triangleq \|I_{c_i}\| \qquad c_i \in C = \{c_1, c_2, \ldots, c_N\} \tag{4}$$

It can be normalised by the total number of pixels n to reflect the probability that a randomly chosen pixel will have the colour c_i. $h(c_i) = H(c_i)/n$.
Hu moments $\mathbf{M}$ is a 7 element feature vector which includes algebraic combinations of normalised central moments [4].
Relative shape measurements $\mathbf{R}$ is a 6 element vector containing the measurements below: Let S denote a stained object, S_b a foreground region (cell) in which S is contained, $a(S), p(S)$ are the area and the perimeter of S, $m(S)$ moment of inertia of S, A_μ average cell area in I. Then $\mathbf{R}$ can be expressed as:

$$\mathbf{R} \triangleq \left\{a(S)/a(S_b), a(S)/A_\mu,\ a(S_b)/A_\mu,\ m(S)/{A_\mu}^2,\ (4\,\pi\,a(S)/p^2(S),\ (4\,\pi\,a(S)/p^2(S_b)\right\} \tag{5}$$

Correlogram $\mathbf{C}$: Suppose, the input image I is quantised to have N distinct colours ($c_i \in Q = \{c_1, c_2, \ldots, c_N\}$) and D denote a predetermined distance set ($k_i \in D = \{k_1, k_2, \ldots, k_M\}$). Then $\mathbf{C}$ is the count of co-occurrences of pixels p_1, p_2 of colour c_i which are k_i distance apart in I:

$$C(c_i, k_i) \triangleq \|I(p_1) = c_i, |p_2 - p_1 = k|\, I(p_2) = c_i\| \tag{6}$$

It can be normalised to reflect the probability that a randomly chosen pixel will have colour c_i and co-occur with another c_i colour pixel in k distance by dividing to $(8k\ h(c_i))$ which is the total number of pixels in distance k times $h(c_i)$ the probability of observing colour c_i. The $8k$ is due to the distance calculation with the city block measure. For example, a centre pixel has 8 related pixels in distance $k = 1$, 16 related pixels in distance $k = 2$.

4.2 Classifier

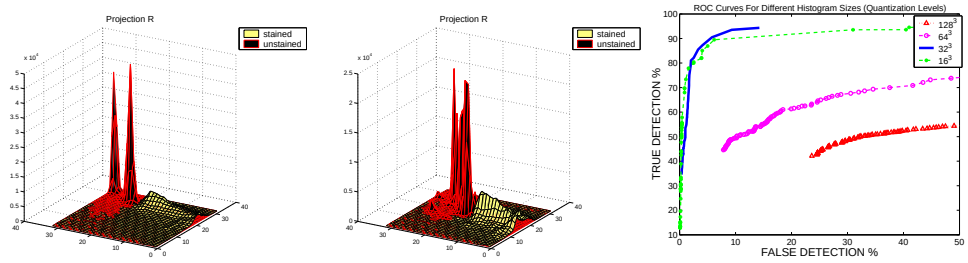
To classify the stained objects we have implemented a distance weighted K-nearest neighbour classifier (Knn-d) [2], [3].
Knn-d: Assigns a query vector to the class $w_c \in W = \{w_1, \ldots, w_m\}$ which has the largest distance weighted majority (number/total distance) in the set of K closest (determined by a distance metric) neighbours in the training set Tr.
Distance Metric: The choice of distance metric is critical for the Knn classifiers. Possible choices for a simple metric can be the L_1 norm, the relative distance metric (RL_1) based on the L_1 norm [5], and L_2 norm. The L_2 norm (Euclidean distance) is known to be less robust against outliers than the L_1 norm (city block distance) [2]. We have chosen to use RL_1 which weights the contribution of each element to the L_1 norm according to the magnitudes and is reported to be more successful in histogram comparisons [5].
Classification Performance: Let $N_t^{w_1}$, $N_t^{w_2}$ show the number of test vectors in set Tt which are known to belong to the w_1 and w_2 classes, respectively. For a diagnosis problem like this, the analysis of four ratios can represent our method's performance. Let tp^{w_i}, fn^{w_i} show the number of correct and missed classifications for class w_i respectively.

$$SE = \frac{tp^{w_1}}{N_t^{w_1}}, \qquad SP = \frac{tp^{w_2}}{N_t^{w_2}}, \qquad PvP = \frac{tp^{w_1}}{tp^{w_1} + fn^{w_2}}, \qquad PvN = \frac{tp^{w_2}}{tp^{w_2} + fn^{w_1}}, \tag{7}$$

where SE, SP, PvP, PvN denote sensitivity, specificity, prediction value positive, and prediction value negative respectively. The former two are usually named as the true detection rates for w_1 and w_2.

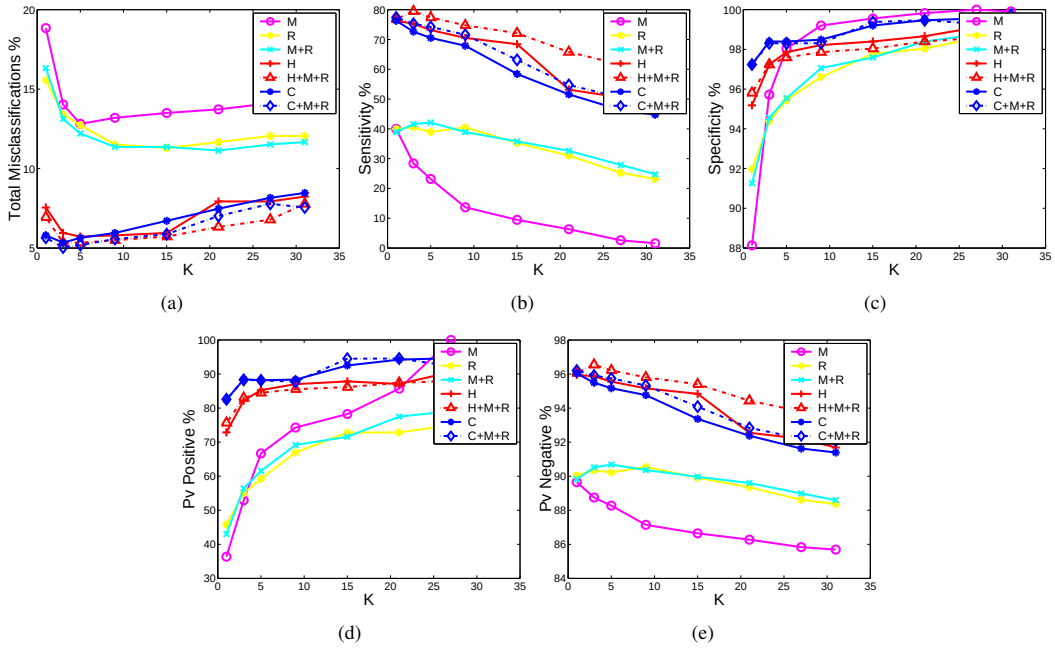
5 Experiments and Results

Stained/Non-Stained Pixel Classification: We have formed 3 separate sets for training (T_a), validation (T_θ), and testing (T_t) containing $286041/37050319$, $205996/22705363$, $168483/25209681$ stained/non-stained pixels respectively. The



(a) (b) (c)

Figure 3. (a) Red axis projections of 3d histograms for stained (light-yellow) and non-stained (dark-red) classes, (b) Same plot in (a) using unnormalised images, (c) ROC curves for different histogram bin sizes.



(a) (b) (c)

(d) (e)

Figure 4. Percentages plotted against K: (a) Total misclassification, (b) SE sensitivity, (c) SP specificity, (d) PvP predictive value of parasites, (e) PvN predictive value of non-parasites. Features H:histogram, M:Hu moments, R:Relative shape measurements, C: Correlogram. (+) denotes concatenation.

two variables (bin size and threshold value) are usually determined by plotting the Receiver Operating Characteristics (ROC): represents the trade off between the true and false detection on a validation test [16]. Then the variables that maximise the area under the ROC curve are chosen for the final tests. Figure 3(c) shows different ROC curves (in T_θ) of different sizes of histograms (quantisation levels :$128, 64, 32, 16$). Examining the associated ROC curves, the most efficient histogram bin size is 32, having the largest area under the curve. In the 32 histogram bin size $\theta = 2.6$ value provides the best rates. Hence, this value was used on the T_t which produced satisfactory results: 88.5%, 5.6% for true and false detection rates respectively. The same experiments were repeated without colour normalisation: the results were 84.5% and 17% respectively. The results show the colour normalisation improves the performance of the Bayesian pixel classification. Cumulative histograms (projections of the 3-d (r,g,b) histograms on the red (r) axis) were plotted in Figure 3(a),3(b) to see the effect of colour normalisation on the distribution of the classes.

Parasite/Non-Parasite Classification: Due to the low number of parasites versus non-parasites in total of 260 images, we have formed 2 separate sets of stained objects for training (T_a), and testing (T_t) containing $175/1312$, $202/1311$, parasite/non-parasite objects respectively. In the T_a, the RGB images were indexed with minimum variance quantisation to have 32 distinct colours, the same indexed colour map is applied to the images in T_t. In the correlogram calculations the distance set was $D = \{1, 2, \ldots, 8\}$ which results in $32x8$ dimensional feature vector. The average cell area A_μ was calculated for every image. All the features were normalised to have zero mean and unit variance.

Figure 4 shows the total misclassification percentage together with SE, SP, PvP and PvN measurements plotted

against increasing K in Knn-d classifier. The results can be regarded as a validation of the overall process and performance comparison of the different features. An examination of the total classification error (Figure 4(a)) indicates the most successful feature to be the $C + M + R$ followed in the order by $H + M + R$, C, and H. However, the difference is not significant considering the cost of calculation time for C. The performances from M and R are low which suggests the colour information is essential. However, they provide slight boosts for the H and C features. To interpret the results in more detail, if we choose the value $K = 3$ in the feature ($C + M + R$), the SE (74%) performance value reveals the probability of the result being positive given that the stained object is a parasite. The SP (98%) value reveals that the probability of the result being negative given that the object is not a parasite. The PvP (88%) value reveals that the probability of the stained object being a parasite given a positive result. The PvN value shows that the negative case (of PvP) is 95%. If the same tests were performed in the other studies which proposed stained object detection as the parasite detection mechanism [10], [11], [12] the performance values would be $SE = 100\%, SP = 0\%, PvP = 15\%, PvN = 0\%$.

6 Conclusion and Discussion

We have described a method to detect malaria parasites in images acquired from Giemsa-stained peripheral blood samples under conventional light microscopes. We have utilised a colour normalisation method to maintain illumination and colour constancy. We have demonstrated a solution for the stain extraction problem with a two class (stained/non-stained) Bayesian classification. The results show that the proposed stained pixel classification performs satisfactorily. The detected stained pixels are further processed to extract features H, M, R, C which are used in the parasite/non-parasite classifier. A knn-d classifier is implemented and detailed evaluations demonstrate individual and concatenated feature performances. Shape features: M and R did not have a significant outcome. Concatenations of $M + R$ were slightly boosting the C and H features. The feature C was more successful than the feature H. However, the difference was not significant which may be caused by the high dimensionality of the feature C.

We have also shown that the stained pixel detection or stained object extraction as proposed in [10], [11], [12] does not lead to parasite detection. Our parasite/non-parasite classifier achieves 74% sensitivity, 98% specificity, 88% positive prediction value, 95% negative prediction value rates when evaluated based on a single observation. However, in a real diagnosis scenario a blood slide from a test case could provide thousands of stained objects. Thus, the diagnostic decision can be made according to the decisions on the total number instead of a single one. To improve the efficiency of the method as a viable malaria diagnosis tool and to reveal the actual diagnosis performance more controlled experiments should be performed and compared to expert manual diagnosis.

References

1. P. Chang and J. Krumm. Object recognition with color cooccurance histograms. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, June 1999.
2. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, New York, 2000.
3. S. Dudani. The distance-weighted k-nearest-neighbor rule. *IEEE Trans. on Systems, Man and Cybernetics*, 6:325–327, 1976.
4. M-K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, 8:179–187, 1962.
5. J. Huang, S. R. Kumar, M. Mitra, W-J. Zhu, and R. Zabih. Spatial color indexing and applications. *Int. Journal of Computer Vision*, 35(3):245–268, 1999.
6. M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *Int. J. Comput. Vision*, 46(1):81–96, 2002.
7. O. Lezoray and H. Cardot. Bayesian marker extraction for color watershed in segmenting microscopic images. In *Proc. Pattern Recognition*, volume 1, pages 739–742, 2002.
8. K. Mitiku, G. Mengistu, and B. Gelaw. The reliability of blood film examination for malaria at the peripheral health unit. *Ethiop.J.Health Dev*, 17(3):197–204, 2003.
9. World Health Organization. What is malaria? fact sheet no94. *http://www.who.int/mediacentre/factsheets/fs094/en/*.
10. KNR. M. Rao. *Application of Mathematical Morphology to Biomedical Image Processing*. PhD thesis, U. Westminster, 2004.
11. C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra. Automatic thresholding of infected blood images using granulometry and regional extrema. In *ICPR*, pages 3445–3448, 2000.
12. C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra. Analysis of infected blood cell images using morphological operators. *IVC*, 20(2):133–146, February 2002.
13. P. Soille. *Morphological Image Analysis*. Springer-Verlag, Heidelberg, 2003.
14. F. B. Tek, A. G. Dempster, and I. Kale. Blood cell segmentation using minimum area watershed and circle radon transformations. In *Proc. Int. Symp. on Mathematical Morphology*, volume 1, pages 739–752, April 2005.
15. F. B. Tek, A. G. Dempster, and I. Kale. A colour normalization method for giemsa-stained blood cell images. In *Proc. Sinyal Isleme ve Iletisim Uygulamalari*, April 2006.
16. A. Webb. *Statistical pattern recognition 2nd Ed.* J Wiley and Sons Inc., New York, USA, 2002.

A Support Vector Machine for 3D texture-based classification of bone in CT images

T. Shepherd [a] and D.C. Alexander [a] *

[a]Centre for Medical Image Computing, University College London, Gower Street, London WC1E 6BT.

Abstract. Bone regions in CT images contain a range of intensities and textures, making their classification a non-trivial task. We show that a Support Vector Machine can be used to distinguish between bone voxels and those from background biological material based on differences in both intensity and texture. The SVM is trained with ground-truth texture samples from several CT images to calculate a decision boundary separating bone from non-bone. Signed distance functions are then calculated for test samples, giving both an estimate of class label and a measure of the classification certainty. Classification performance is compared directly with Linear Discriminant Analysis implemented with information from intensity and local texture measures. Results show that the SVM is a strong discriminator of whole-bone from non-bone. As such we expect the method to be useful in a wide range of image segmentation applications, particularly in combination with more sophisticated segmentation algorithms such as active contours.

1 Introduction

Bone segmentation in CT images typically exploits the high X-ray attenuation of bone material, and for visualisation in 2D, a simple intensity thresholding may suffice in distinguishing bone from background (non-bone) material. However, a single bone varies in density and internal structure, causing image intensity or texture measures to vary widely. Figure (1) shows an axial slice of a CT image of the human ankle, demonstrating the range of intensities and textures and confirming that distributions of intensities from regions of bone and non-bone have significant overlap. The outer bone has apparent gaps due to the variation in density of cortical bone. The inner bone has irregular dark and light patches where calcified tissue intrudes spongy cancellous bone. These issues make segmentation of individual bones of the hands and feet a challenging problem. The segmentation of small joint bones in the wrists and ankles is necessary for modeling joint loading and kinematics [1]. Other applications include the identification of fractured bones in orthopedic surgery planning [2] and quantitative osteoporosis studies [3], where variable mineral density is a fundamental problem.

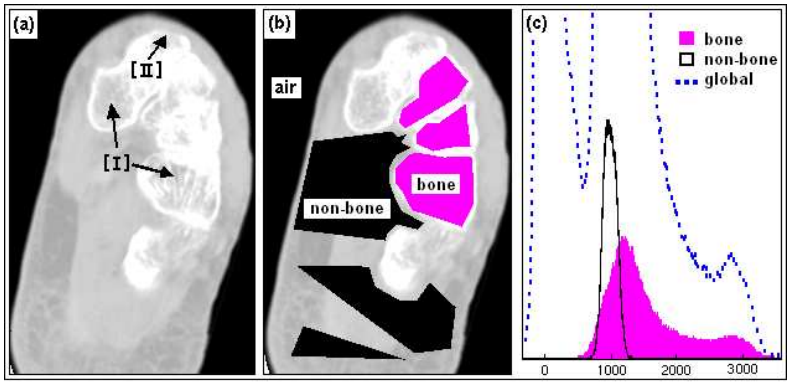


Figure 1. (a) Axial slice of a CT ankle. Arrows highlight examples of [I] the abundance of low intensity cancellous bone in inner-regions and [II] the variability of cortical bone thickness giving apparent discontinuity of the outer-bone surface. (b) Polygons manually drawn to define ground-truth regions of bone and non-bone. (c) Normalised histograms of CT intensity sampled from bone and non-bone regions across the whole of a 3D image. Intensities are in shifted Hounsfield Units as described in section (3). The global histogram is also shown, which includes voxels from air regions at the lowest intensities.

Bone segmentation methods such as active contour models (ACMs) exploit the intensity difference between bone and non-bone, but are confounded in regions of small joint bones where variability has a relatively large effect on the bone's form. Recent developments of ACMs for this application include the coupling of deformable models with region growing and region competition [1] and the formulation of new region energy terms [4]. In this paper, we present

*E-mail: [T.Shepherd, D.Alexander]@cs.ucl.ac.uk

a supervised learning approach to 3D texture classification, which could similarly aid segmentation by deformable contour models. We use a Support Vector Machine (SVM) to characterise the distribution of 3D textures and intensities in each tissue type without the computation of local statistics. We hypothesise that a SVM operating on windows of raw intensities is able to classify bone better than statistical texture measures calculated over the same windows. We compare direct voxelwise classification performance using the SVM with that from using standard texture measures in classical Linear Discriminant Analysis (LDA).

2 Theory

2.1 Linear Discriminant Analysis

In LDA a linear decision boundary is calculated that separates two classes in the same space as the input feature vectors [5]. The training phase in LDA calculates a linear discriminant function

$$g(\mathrm{x}) = \sum_{i=1}^{d} w_i x_i, \tag{1}$$

i.e. a weighted sum of the components of the d-dimensional feature vector $\mathrm{x} = (x_1 ... x_d)^T$. A hyperplane is defined by constant $g(\mathrm{x})$. The task is to find the set of weights $\mathrm{w} = (w_1 ... w_d)$ and an offset from the origin in feature space, of the hyperplane which optimally separates two classes. After training, test data are classified by which side of this decision boundary they lie and a posterior probability of class affiliation is calculated from the perpendicular distance in feature space to the boundary.

2.2 The Support Vector Machine

The support vector method of machine learning was introduced by Cortes and Vapnik in 1995 [6]. A SVM maps input patterns into a higher dimensional space, using a kernel function, where it constructs a decision hyperplane. Maximally separated margins parallel to the hyperplane divide the new feature space into class-specific sub-spaces according to labelled training patterns. Often mapped patterns are not linearly separable, so that no hyperplane can separate two classes perfectly. In this case a cost is assigned to vectors lying between the hyperplane margins which, along with the positioning of margins, controls the optimization. A SVM is hence constructed by defining this cost, its input feature space and the kernel function. The results of training a SVM are the Lagrange multipliers used in optimisation, the data which lie on the hyperplane margins (support vectors) and the perpendicular distance of the hyperplane from the origin of the mapped feature space. New data is then classified by assigning to each pattern a decision-value d_{SVM} which is the perpendicular distance to the hyperplane. By convention the origin of the feature space is shifted to lie on the hyperplane, giving positive and negative decision-values so that sign indicates the predicted class.

2.3 Application to bone texture classification

We use 27-dimensional input patterns, comprising a list of intensities taken from each location in a $3 \times 3 \times 3$ cubic neighbourhood centred on a root voxel, as the input to a SVM. Figure (2) shows example intensity patterns from bone and non-bone regions and demonstrates that the intensity in bone regions varies more widely and the patterns themselves vary more in shape for bone regions than non-bone. Any recurring 3D texture captured by the cubic

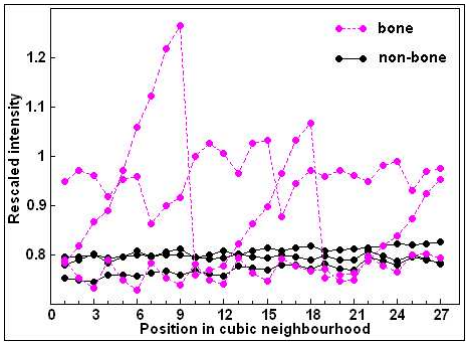


Figure 2. Three examples of 27-intensity patterns chosen at random from each of bone and non-bone regions. Intensity units are rescaled as described in section (4.3). Points are joined to show which intensities belong to a common pattern.

neighbourhood gives its own 'cluster' in input space. The number of clusters reflects the number of distinct textures within a region. By ordering the list of intensities by their position in the neighbourhood, this input feature space respects the spatial relationship between intensities in a neighbourhood, while such texture information is lost in the calculation of local statistics. We calculate four statistical texture measures which, along with CT intensity, give five features. We use the established texture features of local standard deviation, skewness, kurtosis and entropy [7, 8]. These are calculated over a $3 \times 3 \times 3$ cubic neighbourhood, centred on root voxels. LDA classification assigns two probabilities p_+ and p_- to each root voxel, i.e. the probability of belonging to each class. For convenience we convert these to a single signed decision-value p, using the formula:

$$d_{LDA} = p_+ - 0.5 + \left(0.5 \times sign[p_+ - 0.5]\right), \tag{2}$$

which allows the inferred class to be recognised by its sign and facilitates performance evaluation in the same framework as the SVM.

3 Data preparation

We use six CT images of human ankles. All images have in-slice pixel dimensions of 0.4 mm and original slice separation of 2 mm. We interpolate between slices to give 1 mm pseudo-slice separation. Intensities were originally in Hounsfield Units (HU), having a minimum at around -1000, but have been shifted into the positive domain as is common among the CT community. In addition we standardised the images to occupy a common dynamic range before the linear rescaling described in section (4.3). Each image histogram is given the same lower and upper limits. All values below a lower limit I_{low} represent air voxels that do not appear in training or testing samples and were set to this lower limit. The average maximum value over the six images was chosen as an upper limit I_{high}. All values above this value are reduced to I_{high} and for any image with a maximum below this value, the maximum value is increased to I_{high}. Any neighbourhoods containing data at this upper limit are not used in training or testing samples. By imposing these limits we make the method invariant to the presence of both air and highly attenuating foreign bodies such as metallic implants, should we wish to include new data sets in the future.

4 Methods

4.1 Classification and performance evaluation

We use a data analysis framework written in Matlab, which incorporates freely available SVM and LDA tools [9, 10]. Ground-truth locations are stored as co-ordinate arrays built up by drawing polygons on every slice as in figure (1 b). The polygons aim to collect enough examples of all the textures and intensities that belong to each tissue type, and are not required to capture the shape of a given region. To classify unseen test data, we threshold the signed distance to the separating hyperplane. To assess the performance of each classification technique, we compute ROC curves by varying the threshold on d_{SVM} and d_{LDA} and counting true and false positives compared to the ground-truth. The area under the ROC curve provides a measure of the degree of discrimination between texture classes from each method. This area A_{ROC} is unity for a perfect classification and has a lower limit of 0.5 for a complete failure to classify better than random label assignment. We also calculate the operator-point, where a threshold gives optimal classification for specified relative costs associated with false positives and false negatives. This is calculated for use in simple threshold segmentation.

4.2 Experiments

To test each classifier we label ground-truth regions in all six data sets and use a leave-one-out strategy to compute average ROC areas. For a chosen number of data, a random subset of the available ground-truths is taken from the test image, and one-fifth of this number are taken at random from each of the five training images. Permuting the configuration of training and testing images gives six ROC curves. We calculate the mean and standard deviation of classification performance from these six curves. To address the hypothesis that the SVM using raw intensities out-performs standard texture measures used in LDA, we use both methods to classify texture at the same ground-truth locations. To separate the comparison of classification techniques with that of the features used, we use each method with both windowed intensities and statistical texture measures. When using SVM and LDA with texture measures, we use the complete 5-dimensional feature vector. In the case of LDA we also perform separate classifications using 2-dimensional vectors of intensity plus each of the texture measures in turn. This gives insight into the relative discriminatory power of each feature.

4.3 Optimising the SVM

The SVM has various parameters that can be changed and optimised. We chose the Radial Basis Function (RBF) kernel for feature space mapping, which has a parameter γ that controls the kernel width. The cost C assigned to vectors lying between the hyperplane margins, is another flexible parameter. We used a grid-search algorithm to optimise the combination of parameters γ and C. In a scheme similar to that suggested by Hsu et al [11] we identified a local peak in classification performance on the 2D surface of γ and C pairs, at $\gamma = 8, C = 8$. The number of ground-truth patterns used in training was seen to affect classifier performance for numbers less than a few hundred, with little to be gained by using more than a few thousand. We use 5000 training patterns, and classify the same number of test data. Note that this means 1000 samples from each of the five training images in any one configuration. A SVM performs better on data that are rescaled to a small, decimal range [11]. After standardising to a common range we rescale this range to values between 0 and 2, as gave the best performance in trials.

5 Results and discussion

Table (1) gives the results of both classifiers used with the various combinations of intensity and texture features.

Classification Method	**Mean A_{ROC}**	**Standard Deviation**
27-intensity SVM	0.974 (i)	0.0129
5-feature SVM	0.961 (ii)	0.0138
27-intensity LDA	0.960 (iii)	0.0230
5-feature LDA	0.966 (iv)	0.0091
2-feature LDA combining intensity with : std. deviation	0.969 (v)	0.0173
: skewness	0.958 (vi)	0.0205
: kurtosis	0.952 (vii)	0.0267
: entropy	0.959 (viii)	0.0123

Table 1. Classification performances of the SVM and LDA

The 27-intensity SVM gives the highest ROC area of 0.974 (i). However there is little difference between all performances and this classification accuracy is not significantly better ($t = 0.52$) than that of the next best method, namely the 2-feature LDA using intensity and standard deviation (v). Comparison of (i) and (iii) indicates that the technical differences between SVM and LDA, i.e. non-linear mapping and boundary-margin maximisation, improve the separability of intensity-patterns. Interestingly, the SVM technique does not aid the classification of statistical texture features, as indicated by (ii) and (iv). Of all methods using statistical texture measures, the LDA that uses intensity and local standard deviation gives the best classification accuracy (v). The remaining statistical features are the poorest measures of texture in bone and non-bone (vi)→(viii) and do not appear to complement standard deviation information in the LDA, although the difference between (iv) and (v) is not significant and may be due to random fluctuations.

The performance metric A_{ROC} only reflects the proportion of misclassification within a subset of data (ground-truth locations). For more qualitative results, we classify all non-air voxels throughout a 3D dataset and create binary images. Figure (3) shows an axial slice of a 3D image segmented by thresholding the results of some of the classifiers. The threshold chosen gives approximately equal numbers of false positives and false negatives in separate classifications of 5000 ground-truth samples taken from the same image. The 3D image segmented in figure (3) includes slices at the cuneiform region where bones are small and close together. The classification of ground-truth from this image gives ROC areas of 0.957, 0.954 and 0.938 for the 27-intensity SVM, the 2-feature LDA and 5-feature LDA respectively. One advantage of the SVM is revealed by less omission of internal bone in (a) than in (b) or (c). Morphological operations may improve this classification but would not help a more sophisticated segmentation algorithm. Both SVM and 2-feature LDA misclassify data in the narrow regions between close bones and at the skin/fat layer. These sub-classes of non-bone material may have relatively homogenous texture, like cortical bone. However, these regions were not well represented in training samples. Including some examples in the training set should help eliminate these false positives. Panel (c) indicates that, although the inclusion of skewness, kurtosis and entropy may reduce the separability of feature space, (reflected in the noisier background) this information enables the LDA to identify the skin/fat layer as non-bone. The binary nature of the SVM and LDA prohibits us from distinguishing more than 2 tissue types. However, active region models inherently perform binary classifications, so the SVM or LDA would be well suited for use in this type of segmentation [12, 13]. Remaining limitations of the SVM are the relative time complexity and training requirements. We found that a typical classification of 27-intensity or 2-feature vectors took around 40 times longer by the SVM method than LDA. The SVM method also requires larger training sets in order to infer rotation invariance.

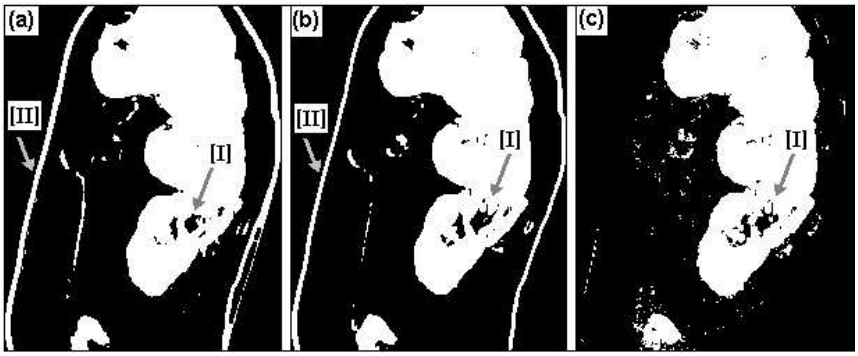


Figure 3. Slice of a 3D image showing bone (white) and non-bone (black) segmented by (a) the 27-intensity SVM, (b) the LDA that uses intensity and standard deviation and (c) the 5-feature LDA. Arrows show the misclassification of [I] internal bone regions and [II] the skin/fat layer.

However, the rotation invariant features used by the LDA do not appear to give this method an advantage for sample sizes above a few thousand.

6 Conclusions

The results show that a SVM can classify bone in CT images based on their varied 3D texture properties. The SVM groups many textures together in the same class, which enables more of the internal bone to be correctly classified. The kernel-based approach conserves the spatial relationship between intensities and removes the need for calculation of local texture features. The SVM gives a slight but not significant improvement over the traditional methods and both SVM and a LDA using statistical texture measures may be incorporated into more sophisticated segmentation algorithms. We intend to use the SVM to drive an active region model, which positions a contour based not only on edge properties but also by maximising a region measure calculated over the data it encloses.

Acknowledgements

This work was funded by an EPSRC studentship. We thank S.J.D. Prince for early reviews of the paper, S.Nedjati-Gilani for a practical introduction to SVMs, Depuy International for providing the images and M.Kiefte for correspondence regarding the use of the LDA toolbox.

References

1. T. Sebastian, H. Tek, J. Crisco et al. "Segmentation of carpal bones from 3d CT images using skeletally coupled deformable models." In *Proceedings of MICCAI'98: First International Conference*, pp. 1184–1194. Lecture Notes in Computer Science, Springer, Berlin, 1998.
2. S. Hirano & Y. Hata. "Fuzzy expert system for foot CT image segmentation." *Image and Vision Computing* **19**, pp. 207–216, 2001.
3. J. Sotoca, J. Iñesta & M. Belmonte. "Hand bone segmentation in radioabsorptiometry images for computerised bone mass assessment." *Computerized Medical Imaging and Graphics* **27**, pp. 459–467, 2003.
4. X. Pardo, M. Carreira, A. Mosquera et al. "A snake for CT image segmentation integrating region and edge information." *Image and Vision Computing* **19**, pp. 461–475, 2001.
5. R. Duda, P. Hart & D. Stork. *Pattern Classification (Second Edition)*. Wiley, *, 2001.
6. C. Cortes & V. Vapnik. "Support-vector networks." *Machine Learning* **20**, pp. 273–297, 1995.
7. T. Southard & K. Southard. "Detection of simulated osteoporosis in maxillae using radiographic texture analysis." *IEEE Transactions on Biomedical Engineering* **43**, pp. 123–132, 1996.
8. R. Haralick. "Statistical and structural approaches to texture." In *Proceedings of IEEE*, pp. 786–804. IEEE, 1979.
9. J. Ma, Y. Zhao, S. Ahalt et al. *OSU-SVM*. Support Vector Machine Toolbox, www.sourceforge.net/projects/svm.
10. M. Kiefte. *DISCRIM*. Linear Discriminant Analysis Toolbox, www.mathworks.com/matlabcentral/fileexchange.
11. C. Hsu, C. Chang & C. Lin. *A practical guide to support vector classification*. Technical Report, National Taiwan University, http://www.csie.ntu.edu.tw/ cjlin/papers/guide/guide.pdf, 2003.
12. J. Ivins & J. Porril. "Statistical snakes: Active region models." In *Proceedings of British Machine Vision Conference*, pp. 377–386. BMVA, 1994.
13. D. Alexander & B. Buxton. "Statistical modeling of colour data." *Int. Journal of Computer Vision* **44**, pp. 87–109, 2001.

Information-Theoretic Unification of Groupwise Non-Rigid Registration and Model Building.

Carole J. Twining[a], T.F. Cootes[a], S. Marsland[b], V. Petrovic[a], R. Schestowitz[a], and C.J. Taylor[a]

[a]Imaging Science and Biomedical Engineering (ISBE), University of Manchester, Manchester, U.K.
[b]Institute of Information Sciences, Massey University, Private Bag 11222, Palmerston North, New Zealand.

Abstract. There is a feature common to both non-rigid registration of a group of images and building a model of a group of images: a dense, consistent correspondence across the group. The former aims to find such a correspondence, whilst the latter requires it. This paper presents the theoretical framework required to unify these two areas, providing a groupwise registration algorithm, where the inherently groupwise model of the image data becomes an integral part of the registration process. The performance of this algorithm is evaluated by using the concepts of generalisability and specificity, which provide an independent metric for comparing various registration algorithms. Experimental results on MR data of brains for various pairwise and groupwise registration algorithms is presented, and demonstrates the feasibility of the combined registration/modelling framework, as well as providing quantitative evidence for the superiority of groupwise approaches to registration.

1 Introduction

Non-rigid registration (NRR) is being increasingly used as a basis for medical image analysis, with applications that include structural analysis, atlas matching and change analysis. There are well-established methods for pairwise image registration(for a review, see e.g., [1]), but often it is necessary to register a group of images. This can be achieved by repeatedly applying pairwise registration, but there is no guarantee that the solution is unique – depending on the choice of reference image, representation of warp, and optimisation strategy, many different results can be obtained for the same set of images. Clearly, this does not form a satisfactory basis for analysis. Our approach is to consider NRR and modelling a group of images [3] as complementary problems: the aim of NRR is to find a meaningful dense correspondence across the group, whereas modelling requires it. Building on the optimal shape model approach of Davies et al [4], we define a minimum description length (MDL) criterion for image model quality, and show that a unique groupwise correspondence can be defined by explicit minimisation of this criterion. NRR and modelling have been combined previously [6], however this required an initial manual labelling of every image. As regards groupwise non-rigid registration, several authors have considered the problem of choosing the best reference image(e.g., [2, 5]). These approaches involve defining a series of independant criteria for what constitutes image matching, how image deformation is weighted against spatial deformation and so on. The advantage of our approach is that we use a single criterion – minimum description length – which can in principle determine not just the groupwise correspondence across the set of images, but also the optimal spatial reference frame, the optimal reference image and, potentially, the optimal model parameters (e.g., number of modes of the model retained). It hence unifies registration and modelling within a *single* coherent theoretical framework.

In this paper, we present a brief description of our framework for groupwise registration, including the MDL objective function, the method of optimisation, and metrics for evaluating different groupwise correspondences. We evaluate the performance of a range of pairwise and groupwise approaches to registering a set of brain images, and show that the groupwise approach gives quantitatively better performance than pairwise.

2 Spatial & Pixel/Voxel-Value Transformations

The correspondence across the group has to be consistent, and one way to ensure this is to define all correspondences w.r.t. a spatial reference frame. We hence define the following basic notational conventions , taking as our example the simplest case of a spatial warp directly between a training image frame and a reference frame (see Fig. 1):

- X_0 is the regular grid of pixel/voxel positions on which each of our images is defined.
- $\mathcal{R}$ is the spatial frame of the reference. A reference image $I_{\mathcal{R}}(X_0)$ is the values of a function $I_{\mathcal{R}}$ at the positions X_0.
- The set of N training images $\{I_{\mathcal{T}_i}(X_0) : i = 1, \ldots N\}$. Image i, function $I_{\mathcal{T}_i}$ in spatial frame $\mathcal{T}_i$.

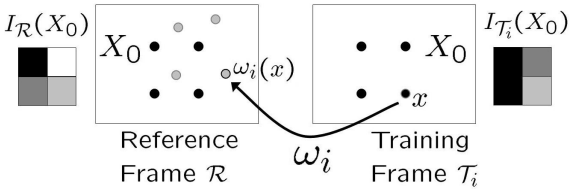


Figure 1: A spatial warp ω_i from training frame $\mathcal{T}_i$ to reference frame $\mathcal{R}$. X_0 (black filled circles) is the set of regular voxel positions, with the grey filled circles being the warped voxel positions $\omega_i(X_0)$.

The dense correspondence between a training image frame $\mathcal{T}_i$ and the reference frame $\mathcal{R}$ is defined by a spatial warp $\omega_i : x \in \mathcal{T}_i \mapsto \omega_i(x) \in \mathcal{R}$. The warp ω_i also induces a mapping between the function spaces (that is, it warps images between frames). Mathematically, there are two such mappings:

The push-forward: $\omega_i : I_{\mathcal{T}_i} \mapsto I_{\mathcal{T}_i}^{\omega_i} \doteq \omega_i(I_{\mathcal{T}_i}), \qquad I_{\mathcal{T}_i}^{\omega_i}(\omega_i(x)) \doteq I_{\mathcal{T}_i}(x)$

The pullback: $\omega_i^* : I_{\mathcal{R}} \mapsto I_{\mathcal{R}}^* \doteq \omega_i^*(I_{\mathcal{R}}), \qquad I_R^*(x) \doteq I_{\mathcal{R}}(\omega_i(x))$

*carole.twining@manchester.ac.uk

Because of re-sampling issues for regular and irregular pixel grids, the pullback ω_i^* is easier to compute than the push-forward mapping, so that we will use the **pullback** mapping wherever possible, where the direction of flow of image information is in the **opposite** direction to that of the spatial mapping. The ability to map images between frames means we can compare images. We will denote a general discrepancy-image by ΔI. In the example above, a discrepancy image in the frame $\mathcal{T}_i$ is: $\Delta I_{\mathcal{T}_i}(X_0) = I_{\mathcal{T}_i}(X_0) - I^*_{\mathcal{R}}(X_0) \Longrightarrow (\Delta I_{\mathcal{T}_i} \circ \omega_i^*) I_{\mathcal{R}}(X_0) \equiv I_{\mathcal{T}_i}(X_0)$ where $(\Delta I_{\mathcal{T}_i} \circ \omega_i^*)$ is taken to denote the composition of a pullback mapping ω_i^* and a voxel-value deformation $\Delta I_{\mathcal{T}_i}$. The pixel/voxel-value deformation in this case is defined such that when applied to the warped reference image $I^*_{\mathcal{R}}(X_0)$ it exactly recreates the training set image $I_{\mathcal{T}_i}(X_0)$. It is important to note that in general these two classes of transformations **do not** commute. We now have a general class of image deformations, composed of a spatial part and a discrepancy image part – we will denote such a general combined deformation by capital greek letters (e.g., Ω_i). A more complicated situation is shown in Fig. 2. This shows the reference image being transformed into a training image $I_{\mathcal{T}_i}$, by a sequence of two combined transformations Υ_i then Ω_i. We take this approach since, if we are to model combined transformations across the group of images, we need them to be applied in a **common** frame. So, the spatial transformations $\{\upsilon_i\}$ and the discrepancy images $\{\Delta_i I_{\mathcal{R}}\}$ are all applied in the reference frame $\mathcal{R}$, hence can be modelled across the group. The spatial warp ω_i is now just from the training frame $\mathcal{T}_i$ to the intermediate frame $\mathcal{M}_i$, the corresponding combined warp Ω_i being constructed using the pullback ω_i^* and the discrepancy image $\Delta I_{\mathcal{T}_i}$, which is calculated in a manner analogous that given previously, but with the intermediate image $I_{\mathcal{M}_i}$ taking the place of the reference image $I_{\mathcal{R}}$.

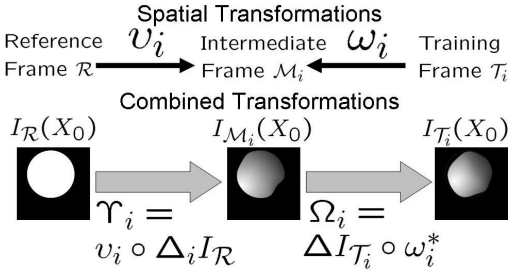


Figure 2: **Top:** The spatial transformations (black arrows) between reference, intermediate and training image frames. **Bottom:** The corresponding combined (spatial and voxel-intensity) transformations (broad grey arrows) between images.

3 The Objective Function

As we explained in the Introduction, we have chosen to define the optimal groupwise non-rigid registration as that which minimises an objective function based on the minimum description length (MDL) principle [8]. The basic idea behind MDL is that we consider transmitting our dataset to a receiver, encoding the dataset using some model. Using the structure and notation defined in the previous section, the data we have to transmit is the reference image $I_{\mathcal{R}}$ and the set of combined deformations $\{\Upsilon_i, \Omega_i\}$ that enable us to exactly reconstruct each training image. The total description length can hence be decomposed thus:

$$\begin{array}{lllll} \mathfrak{L}_{\text{total}} = & \mathfrak{L}_{\mathcal{R}}(\mathcal{R}, I_{\mathcal{R}}) & +\mathfrak{L}_{\text{params}} & +\mathfrak{L}_{\text{group}}(\{\Upsilon_i\}) & +\mathfrak{L}_{\text{residuals}}(\{\Omega_i\}) \\ & \text{Reference frame \& reference image} & \text{Parameters of groupwise model} & \text{Encoded using groupwise model} & \text{Encoded residuals} \end{array} \quad (1)$$

Actual description lengths are computed using the fundamental result of Shannon [10] – if there are a set of possible, discrete events $\{A\}$ with associated encoding-model probabilities $\{p_A\}$, then the optimum code length required to transmit the occurrence of event A is given by:$\mathfrak{L}_A = -\ln p_A$ nats.[1] See [11–13] for details as to the explicit calculation of description lengths.

4 The Algorithmic Framework

4.1 Initialisation

In [11], an algorithm was presented to find an initial correspondence using MDL. The structure of the algorithm followed that shown in Fig. 1. The free variables were the set of spatial warps $\{\omega_i\}$, initialised to the identity $\mathbb{I}$, and the reference image was taken to be the mean of the training images, pulled-back using the inverses $\{\omega_i^{-1}\}$. This algorithm was fully groupwise, in that changes to any of the $\{\omega_i\}$ change the reference, hence change the description length for all of the images in the set. However, the calculation of the inverse warps (or alternatively the push-forward mappings generated by $\{\omega_i\}$) is computationally expensive. We propose here a computationally cheaper initialisation algorithm, within the structure shown in Fig. 2. We keep the idea from the algorithm presented in [11], of initial image estimates based on averages of pushed-forward training images, but instead choose to populate the intermediate images, using the leave-one-out means:

$$I_{\mathcal{M}_i}(X_0) = \frac{1}{N-1} \sum_{j \neq i} [\omega_j^{-1^*}(I_{\mathcal{T}_j})](X_0), \quad (2)$$

with $\{\upsilon_i = \mathbb{I}\}$. We do not explicitly assign a value to the reference image. But we would expect the intermediate images to mutually converge as the algorithm progresses and the images are brought into alignment, so that $\{\Delta_i I_{\mathcal{R}} \mapsto \varnothing\}$.

[1] The **nat** is the analogous unit to the **bit**, but using a base of e rather than base 2.

Algorithm 1 : MDL NRR Initialisation

$\{\omega_i = \mathbb{I},\ i = 1, \ldots N\}$ **%:Initialize warps to the identity.**
Repeat
  Randomize the order of the set of training images $I_{\mathcal{T}_i}(X_0)$, indexed by i.
  For $i = 1$ to N **do**
    Optimise $\mathfrak{L}_{\text{init}}(\{\omega_k\})$ w.r.t. spatial warp ω_i.
    Update Intermediate Images $\{I_{\mathcal{M}_j}(X_0) : j \neq i\}$. **%:Using equation (2).**
  End
Until convergence

The true description length is estimated thus :

$$\mathfrak{L}_{\text{init}}(\{\omega_i\}) = \underbrace{\tfrac{1}{N}\sum_i \mathfrak{L}_{\text{Hist}}(I_{\mathcal{M}_i}(X_0))}_{\text{Estimate of } \mathfrak{L}_{\text{Hist}}(I_{\mathcal{R}}(X_0))} + \underbrace{\sum_i \mathfrak{L}(\omega_i)}_{\text{Spatial Warps}} + \underbrace{\sum_i \mathfrak{L}(\Delta I_{\mathcal{T}_i}(X_0))}_{\text{Discrepancy Images}}. \quad (3)$$

The pseudocode for the initialisation algorithm is given in Alg. 1.

4.2 Groupwise Models

We have shown how to initialise the registration algorithm, we now have to consider the explicit groupwise model. One method would be to build some default generative model of the set of deformations $\{\Upsilon_i\}$, and then search within the space of this model. However, this approach suffers from two drawbacks; firstly, the use of a default model (such as a gaussian) would bias the results, since it would tend to force the deformations to have a gaussian distribution, rather than finding the best deformations. The second drawback is computational – if we alter Υ_i, we have to then re-calculate Ω_i so that the combined deformation does indeed re-create our target training image $I_{\mathcal{T}_i}(X_0)$. This means that we have to re-calculate the intermediate image $I_{\mathcal{M}_i}(X_0)$, which means either calculating a pushforward mapping via v_i, or a pushback via v_i^{-1}, both of which are computationally expensive.

Algorithm 2 : MDL NRR & Groupwise Model Building

Run Algorithm 1 **%:Output is** $\{I_{\mathcal{M}_i}(X_0), \omega_i, \Delta I_{\mathcal{T}_i}(X_0)\}$
$v_i \Leftarrow \mathbb{I}$ **%:Initial Shared frame for all Intermediate Images**
$I_{\mathcal{R}}(X_0) \Leftarrow \frac{1}{N}\sum_i I_{\mathcal{M}_i}(X_0)$ **%:Estimate Reference as Mean**
$\Delta_i I_{\mathcal{R}} \Leftarrow I_{\mathcal{M}_i}(X_0) - I_{\mathcal{R}}(X_0)$ **%:Maintain Intermediate Images**
BUILD & TEST GROUPWISE MODEL OF $\{\Upsilon_i \equiv v_i \circ \Delta_i I_{\mathcal{R}}\}$
$(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, v_i, \omega_i, I_{\mathcal{M}_i}, \Delta I_{\mathcal{T}_I}\}) \Leftarrow$ **TEST-MODEL**$(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, v_i, \omega_i\})$
MAIN LOOP
Repeat
  Repeat
    Randomize the order of the set of training images $I_{\mathcal{T}_i}(X_0)$, indexed by i
    OPTIMISE WARPS ω_i
    For $i = 1$ to N **do**
      Optimise $\mathfrak{L}_{\text{total}}$ w.r.t. spatial warps ω_i. **%:**$\mathfrak{L}_{\text{total}}$ **calculated from eq. (1)**
    End
  Until convergence
  RE-BUILD MODEL
  $(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, v_i, \omega_i, I_{\mathcal{M}_i}, \Delta I_{\mathcal{T}_I}\}) \Leftarrow$ **TEST-MODEL**$(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, v_i, \omega_i\})$
Until convergence

Function TEST-MODEL: BUILD & TEST GROUPWISE MODEL

$\mathfrak{L}_{\text{old}} \Leftarrow \mathfrak{L}_{\text{total}}(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, v_i, \omega_i\})$ **%:Description Length** $\mathfrak{L}$ **before modelling, eq.(1)**
$v_i^{\text{new}} \Leftarrow \omega_i^{-1} \circ v_i$ **%:Put all spatial warp into** v_i
BUILD MODEL
$(I_{\mathcal{R}}^{\text{new}}, \{\Delta_i^{\text{new}} I_{\mathcal{R}}, v_i^{\text{new}}\}) \Leftarrow$ **MODEL**$(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, v_i^{\text{new}}\})$
$\omega_i^{\text{new}} \Leftarrow v_i^{\text{new}} \circ (v_i^{-1} \circ \omega_i)$ **%:Reset** ω_i^{new} **to maintain spatial correspondence**
$\mathfrak{L}_{\text{new}} \Leftarrow \mathfrak{L}_{\text{total}}(I_{\mathcal{R}}^{\text{new}}, \{\Delta_i^{\text{new}} I_{\mathcal{R}}, v_i^{\text{new}}, \omega_i^{\text{new}}\})$ **%:Description Length after modelling**
If $\mathfrak{L}_{\text{new}} \leq \mathfrak{L}_{\text{old}}$ **then**
  $\omega_i \Leftarrow \omega_i^{\text{new}}$, $v_i \Leftarrow v_i^{\text{new}}$, $I_{\mathcal{R}} \Leftarrow I_{\mathcal{R}}^{\text{new}}$, $\Delta_i I_{\mathcal{R}} \Leftarrow \Delta_i^{\text{new}} I_{\mathcal{R}}$ **%:Accept new values**
  $I_{\mathcal{M}_i}(X_0) \Leftarrow (v_i \circ \Delta_i I_{\mathcal{R}}) I_{\mathcal{R}}(X_0)$ **%:Reset Intermediate Images**
  $\Delta I_{\mathcal{T}_i}(X_0) \Leftarrow I_{\mathcal{T}_i}(X_0) - [\omega_i^*(I_{\mathcal{M}_i})](X_0)$ **%:Reset discrepancies in Training frame**
End

We take an alternative approach, which is to optimise the $\{\omega_i\}$. As in Alg. 1, this only involves computing the pullback ω_i^*. So, after we have optimised the set $\{\Omega_i\}$, we then transfer of much of this combined deformation as possible from the intermediate frame $\mathcal{M}_i$ to the equivalent deformation applied in the reference frame $\mathcal{R}$. We can then construct a model in the reference frame. The proposed algorithm is given in Alg. 2. Lines 1-5 are just the initialisation stages, which run the previous initialisation algorithm. The transfer between $\{\Omega_i\}$ and $\{\Upsilon_i\}$ is given in lines 2-3 of the function TEST-MODEL. An important point to note is in line 4 of that function – we maintain the spatial correspondence that we have previously found, despite moving spatial warps between frames. We then build a model of the set of combined deformations $\{\Upsilon_i = (\upsilon_i \circ \Delta_i I_{\mathcal{R}})\}$ and the reference image $I_{\mathcal{R}}(X_0)$. The modelled deformations are not necessarily the same as the input deformations to the modelling process, which is the reason for the resetting in line 5. We then accept this model provided that it decreases the total description length.

5 Implementation Issues

Consider the relation of spatial frames for the groupwise algorithm (e.g., see Fig. 2 and Alg. 2) – it is clear that we require a description of spatial warps $\{\omega_i, \upsilon_i\}$ that allows us to efficiently invert and concatenate warps, as well as a description which allows us to represent a set of warps (i.e.,$\{\upsilon_i\}$) within a common representation for the purposes of modelling. Such a description is provided by spline-based formulations which interpolate the movement of general points from the movement of a set of nodes/knotpoints, where the knotpoints can take **arbitrary** positions. In the experiments which follow, we use both the clamped-plate spline, and an efficient spline based on the piecewise-linear interpolation of movements across a tesselated set of knotpoints in either 2D or 3D. Successive optimisations of the set $\{\omega_i\}$ in Alg. 2 are calculated by adding knotpoints to the previously-optimised set (hence increasing the resolution of the spatial warp). These knotpoints are also chosen in a data-driven manner (e.g., image features such as edges, or places of high discrepancy – see [7, 9] for further examples of such data-driven techniques). The optimisation scheme for the knotpoints is a simple gradient descent – points are moved singly to estimate the gradient direction for the objective function, but moved all at once using a line search.

6 Model Evaluation Criteria

In order to compare different algorithms for non-rigid registration and model building, we need to have some quantitative measures of the properties of a given model. Following Davies et al. [4], we use two measures of model performance:

Generalisability: the ability to represent unseen images which belong to the same class as images in the training set.

Specificity: the ability to only represent images similar to those seen in the training set.

Let $\{I_{\mathfrak{a}}(X_0) : \mathfrak{a} = 1, \ldots \mathfrak{N}\}$ be some large set of images, generated by the groupwise model, and having a distribution which is the model distribution. We then define the following quantitative measures:

Generalisability: $G = \frac{1}{N} \sum_{i=1}^{N} \min_{\text{w.r.t. } \mathfrak{a}} \left(|I_{\mathcal{T}_i}(X_0) - I_{\mathfrak{a}}(X_0)| \right)$, **Specificity:** $S = \frac{1}{\mathfrak{N}} \sum_{\mathfrak{a}=1}^{\mathfrak{N}} \min_{\text{w.r.t. } i} \left(|I_{\mathcal{T}_i}(X_0) - I_{\mathfrak{a}}(X_0)| \right)$

where the distance $|\cdot|$ is a measure of the distance between two images, such as the Euclidean or shuffle distance.

7 Experiments: Evaluation of Pairwise & Groupwise Registration and Models

We have previously performed experiments to validate our MDL objective function and model evaluation criteria, see [13] for details. Here we investigate the performance of several different non-rigid registration methods, including that presented in this paper. Although all the methods we have described can be used in 3D, it was impractical to run the very large set of experiments required in the time available, thus we present results for 2D images of the brain. To evaluate different methods of non-rigid registration we used a dataset consisting of 104 2D MR slices of brains taken from normals; the initial 3D data set was affinely-aligned, and then the corresponding slice extracted from each example. In order to compare different registration strategies, for each technique we registered the entire set of 104 images and built the statistical models of shape and appearance given by the found correspondence, using the nodes/knotpoints used during the registration. We then computed the Generalisability G and Specificity S for each model (generating 1000 model examples in each case, and using a 5-pixels square sample region for the shuffle distance), enabling a quantitative comparison of the registration strategies from which each model was derived. The strategies tested were:

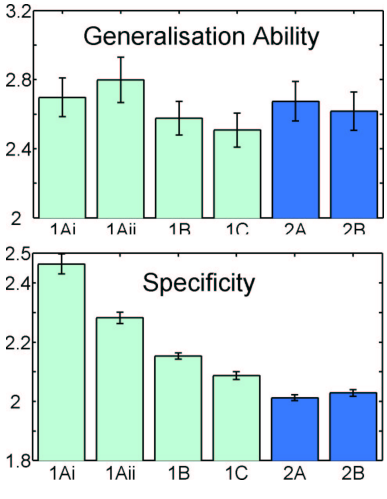


Figure 3: Generalisation ability and Specificity for the strategies listed – dark bars groupwise, light bars pairwise.

1. Pairwise Registration:

A: Image from training set chosen as reference & 16×16 regular grid of nodes:

(i) Residuals calculated in reference frame

(ii) Residuals calculated in training frame

B: As above, but removing points from the grid in regions of low texture variance.
C: Ditto, but moving points to nearby strong edges.

2. Groupwise Registration:

A: Registering to Intermediate Images estimated as the leave-one-out means (Alg. 1).
B: Registering to Intermediate Images estimated using the leave-one-out models.

Note that for **1**, we tried a selection of images from the training set as the reference, and choose that which gave the best results in terms of the evaluation criteria. Strategy **2B** can be viewed as an approximation to the full algorithm given in Alg. 2; in the same way that in the initialisation algorithm (Alg. 1) we estimate the Intermediate Images $\{I_{\mathcal{M}_i}\}$ using the leave-one-out mean, in this case we estimate them by finding the closest fit to the training image $I_{\mathcal{T}_i}$ using the shape model built from all the other examples and the current best estimate of their correspondence. We then optimise the description length of the shape and texture discrepancies between this model estimate and the training image. Note that we do not model the texture at this intermediate stage – this is because in the inner loops of Algs. 1&2, the warps $\{\omega_i\}$ at each spatial resolution are fully optimised, hence can then be modelled, whereas the texture discrepancy is merely continually reduced. The results of this comparison are given in Fig. 3.

8 Discussion & Conclusions

We have presented a principled framework for groupwise non-rigid registration, based on the concept of minimum description length. A groupwise model of shape and appearance is an integral part of the registration algorithm, hence the registration also produces an optimal appearance model. We have given a brief description of a practical implementation of the basic ideas. The key results are those summarised in Fig. 3. These show that our groupwise approach achieves better Specificity than several different pairwise approaches. They also show the importance of measuring errors in the correct frame of reference. Further work is required to implement more sophisticated versions of our groupwise approach, and to provide a more comprehensive set of comparisions to alternative approaches. Our initial results are, however, extremely encouraging.

Acknowledgements: This research was supported by the MIAS IRC project, EPSRC grant No. GR/N14248/01, UK Medical Research Council Grant No. D2025/31 (*"From Medical Images and Signals to Clinical Information"*). S. Marsland was supported by the Marsden Fund grant `MAU0408`, Royal Society of New Zealand, *"A principled approach to the non-rigid registration and structural analysis of groups of medical images"*.

References

1. Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977 – 1000, 2003.
2. K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. *Proceedings of the IEEE Symposium on Biomedical Imaging (ISBI)*, pages 908–911, 2004.
3. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:681–685, 2001.
4. R. H. Davies, C. J. Twining, P. D. Allen, T. F. Cootes, and C. J. Taylor. Shape discrimination in the hippocampus using an MDL model. In *Proceedings of IPMI 2003*, pages 38–50, 2003.
5. B. Davis, P Lorenzen, and S. Joshi. Large deformation minimum mean squared error template estimation for computational anatomy. *Proceedings of the IEEE Symposium on Biomedical Imaging (ISBI)*, pages 173–176, 2004.
6. A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modelling. *IEEE Transactions on Medical Imaging*, 21(9):1151–1166, 2002.
7. S. Marsland and C. J. Twining. Constructing data-driven optimal representations for iterative pairwise non-rigid registration. *Lecture Notes in Computer Science*, 2717:50–60, 2003.
8. J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific Press, 1989.
9. J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D.L.G. Hill, and D. J. Hawkes. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Proceedings of MICCAI 2001*, number 2208 in Lecture Notes in Computer Science, pages 573 – 581, 2001.
10. C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423,623–656, 1948.
11. C. J. Twining, S. Marsland, and C. J. Taylor. Groupwise non-rigid registration: The minimum description length approach. In *Proceedings of the British Machine Vision Conference (BMVC)*, volume 1, pages 417–426, 2004.
12. C.J. Twining, S. Marsland, and C.J. Taylor. A unified information-theoretic approach to the correspondence problem in image registration. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 2004.
13. C.J. Twining, Tim Cootes, Stephen Marsland, Vladimir Petrovic, Roy Schestowitz, and Chris J. Taylor. A Unified Information-Theoretic Approach to Groupwise Non-Rigid Registration and Model Building In *Proceedings of the 19th International Conference on Information Processing in Medical Images (IPMI'05)*, volume 3565 of Lecture Notes in Computer Science, pages 1-14. Springer, 2005.

Distinguishing between Classes of Benign Prostatic Hyperplasia using MR Images

S. P. Ang, P.D. Allen[1], J. Graham and C.E. Hutchinson

Imaging Science and Biomedical Engineering, University of Manchester, UK

Abstract. Benign Prostatic Hyperplasia has two histologically distinct components. It is important to identify which of the two is predominant, as this influences treatment. The two forms can be distinguished radiologically on MR images using a number of visual and semi-quantitative cues. In this paper we investigate the use of texture features based on co-occurrence histograms to provide a quantitative assessment of the tissue type. We show that appropriate texture measures associate strongly with radiological classification based on published criteria.

Introduction

Benign Prostatic Hyperplasia (BPH) is a non-cancerous enlargement of the prostate that affects 80% of men over the age of 40 and 95% of men over 80. The enlargement can cause constriction of the urethra, which runs through the prostate, resulting in obstruction of urinary flow [1]. In 25% of men aged 80, the symptoms are sufficiently severe to require treatment by transurethral resection of the prostate (TURP). However, this treatment has high cost, morbidity (16%) and mortality (2.01%) [2], so alternative treatments are sought.

Anatomically, the prostate can be divided into a number of zones: peripheral, central, transitional and fibro-muscular. In MR images only two regions can be distinguished: the peripheral zone (PZ) and the remaining anatomical zones collectively known as the central gland (CG) (see figure 1). BPH primarily affects the transitional zone, which, in the presence of disease, is the main component of the central gland, so one approach to recognising and estimating the severity of BPH is to measure the ratio of central gland to peripheral zone volumes (CG/PZ). This approach requires differential segmentation of the MR images [3]. However, a different approach is possible involving characterisation of the tissue of the central gland. There are two principal components of BPH: muscular or stromal and glandular (non-stromal) [4]. Ishida et al [5] observed that, in T2-weighted MR images, some areas of the prostate show a variety of heterogeneous and nodular patterns associated with BPH, and other studies have also made qualitative descriptions of enhanced tissue heterogeneity in BPH [6]. These observations may be associated with differing proportions of stromal and glandular tissue present. Measurement of the degree of tissue inhomogeneity may not only provide a useful biomarker to identify the severity of disease, but may permit differential diagnosis, with consequences for pharmaceutical treatment. For example α-blockers may be effective at treating patients with a large percentage of stromal growth, while predominantly glandular BPH may be more appropriately treated with 5-α-reductase inhibitors[4]. Accurate classification into the two types currently requires biopsy followed by histological examination. Ishida et al [5] showed that the histological type of BPH could be accurately determined by visual inspection of T2-weighted MR images using a set of semi-quantitative criteria involving the heterogeneity of the central gland (CG)and the relative volumes of the central gland and peripheral zone (PZ).

We previously showed [7] that texture features, based on the grey-level histogram and granulometry, showed association with visual appearance of the central gland. In this study we show that scale-dependent texture features are strongly associated with radiological classification using the criteria of Ishida et al.[5].

Methods

Image Data

We have used T2-weighted fat suppressed (T2FS) images as these result in good separation of the prostate from most surrounding tissue and enhanced distinction between CG and PZ compared to T2 or T1 weighting. Images from 19 patients with BPH were collected using a Philips Gyroscan Intera 1.5T MR scanner (software version NT8.11, synergy body coil). Each Image consists of 50 axial slices with a thickness of 2mm and in in-plane resolution of 1.0mm. (These images are similar to those used in our earlier study, with slightly higher in-plane resolution.) Each volume image has been manually segmented in 3D to define the surfaces of the whole prostate and the central gland.

[1] Contact: Philip.Allen@manchester.ac.uk

Figure 1 shows two examples of axial slices in which the central gland and peripheral zone have been labelled. Figure 1a shows the inhomogeneous, granular appearance characteristic of predominantly glandular (non-stromal) BPH; the compression of the peripheral zone due to the enlargement of the central gland is very clear in this case. Figure 1b shows the rather smoother appearance indicative of stromal BPH. Ishida's criteria for recognition of glandular BPH are:

- The presence of inhomogeneous high-signal texture
- Ratio of central gland to total prostate volume > 0.75
- The presence of a surgical capsule.

The surgical capsule can be seen in figure 1a as the dark (low signal) border around the central gland, caused by compressed central gland tissue..

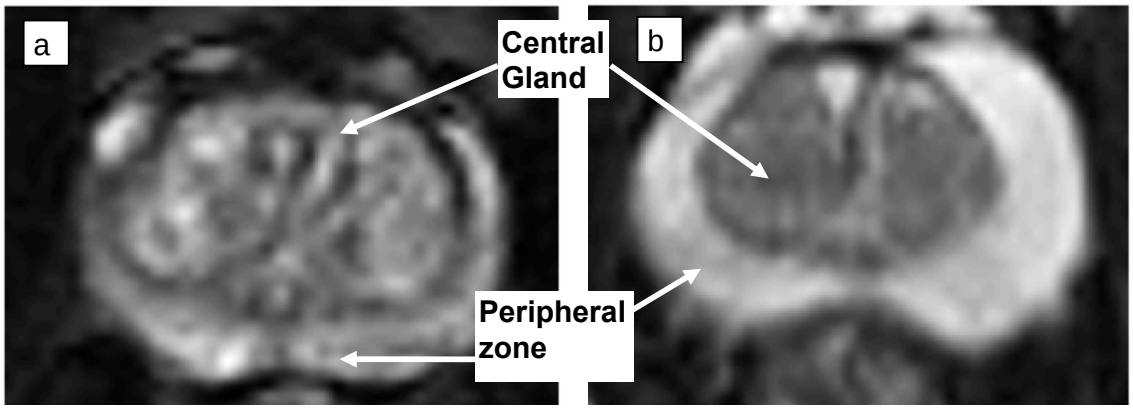


Figure 1 The central axial slice through T2 Fat-suppressed MR images of the prostate showing: (a) the granular inhomogeneous appearance characteristic of the central gland and the compressed peripheral zone associated with glandular (non-stromal) BPH, and (b) the more homogeneous appearance of the central gland and larger peripheral zone in stromal BPH.

Grey-level Co-occurrence Matrices.

Grey-level co-occurrence matrices are a classical approach to texture measurement [8]. The matrix is a histogram of the joint occurrences of grey-levels *i* and *j*. Its formation is illustrated in figure 2. A separation *d* (in distance and direction) is chosen. Each pixel in the image gives rise to a pair of grey levels *(j, j)*, defined by *d*. The matrix is a joint histogram of (*i,j*) pairs. Where the texture is expected to be isotropic (as in this case), grey-level *j* can be taken to be an average grey level on the circumference of a circle of radius *d* (as shown on figure 2). Having built the histogram a number of measures can be calculated to represent different textural properties. Treating the normalised matrix values as probabilities of observing a pair of grey-levels *(j, j)*, we can define

Homogeneity: $\sum_{i,j} \frac{P(i,j)}{1+|i-j|}$, correlation: $\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j}$

contrast: $\sum_{i,j} |i-j|^2 P(i,j)$, energy: $\sum_{i,j} P(i,j)^2$,

A number of other measures corresponding to different aspects of image texture can be defined [8]; contrast, correlation and energy proved to be the most useful in this study.

Each matrix, and hence each derived measure, corresponds to a particular value of the displacement or radius *d*. This parameter is clearly related to the size of texture elements that is to be expected in the images.

For this study we did not choose a particular scale, but constructed a texture signature for each texture measure, calculating the texture feature value for $1 \leq d \leq 20$. Figure 3 shows an example of a texture scale signature corresponding to the contrast measure for one image. In order to use the scale texture signatures for classification, a number of features were used to summarise the signature. These were: the mean value across scales, the weighted

mean (weighted by the scale), the median, the entropy (treating the curve as a distribution and calculating $e = \sum_{i,j} p \log(p)$, and the area under the curve. The last measure calculated the area enclosed by the signature curve and the diagonal line between the first and last elements (see figure 3).

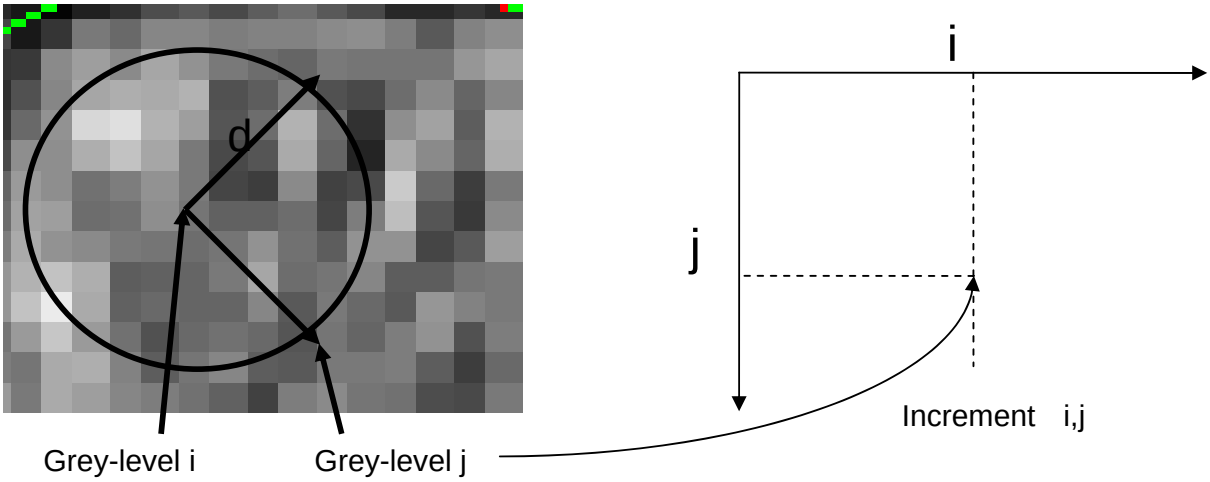


Figure 2. Illustrating the formation of the grey-level co-occurrence matrix. The matrix is a joint histogram, element *i,j* being incremented when a pair of grey levels *(i, j)* is observed at the selected separation *d*. For a homogeneous texture, *j* can be taken to be the average grey level around the circumference of a circle of radius *d*.

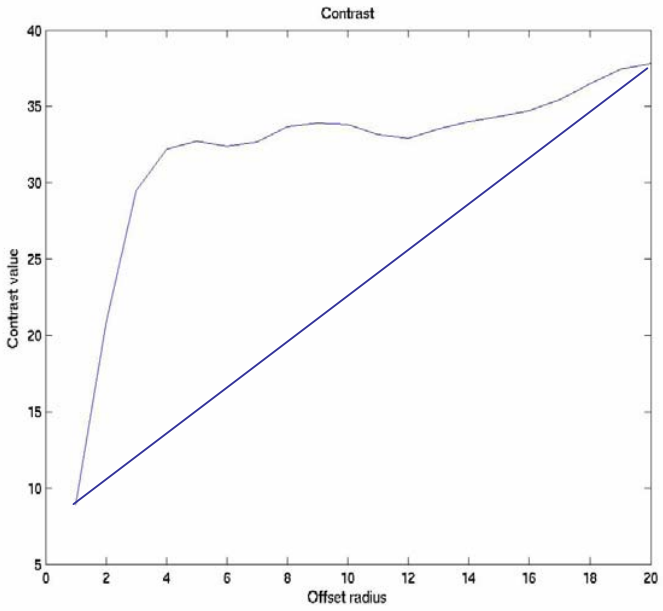


Figure 3. An example texture scale signature, showing the value of contrast derived from the co-occurrence matrix for a particular image as a function of the radius of the sampling circle, *d*. Texture values were calculated for $1 \le d \le 20$.
Features measured from the signatures include mean, weighted mean and median contrast and the area under the curve, i.e. the area enclosed by the signature curve and the diagonal line joining the first and last elements.

Experiments on prostate images

The co-occurrence matrices give texture measures on 2D images. The central gland of the prostate covers several slices, depending on the size of the prostate. In our experiments we measured texture on the central slice of the central gland, and included a number of adjacent slices, the total number of slices being a variable parameter in the experiments. As prostate size is a potentially indicative factor for BPH, we sought to make the texture measure independent of prostate size. We calculated co-occurrence matrices over a circular area of constant size in each

image. This area was chosen to be the size of the smallest prostate in the set. For larger prostates several areas were sampled and the average features calculated.

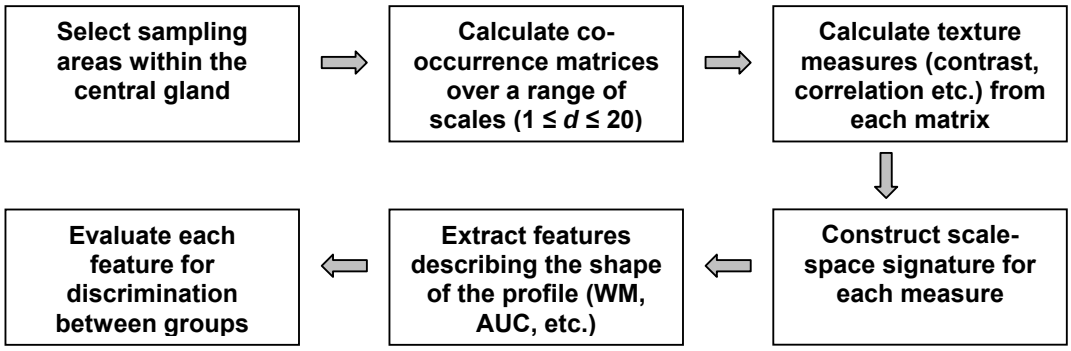


Figure 4. Flow-chart summarising the feature extraction and evaluation procedure..

Figure 4 illustrates the feature extraction process.The images were divided into two groups using the three criteria of Ishida et al. [5]. These groups were tested on whether they were significantly separated on any of the texture measures. The distributions of values were first tested for normality using a Lilliefors test and a t-test or Mann-Whitney test applied according to whether the data were normally distributed or not.

Table 1 shows the texture features that gave the most significant discrimination of the groups

Texture measure	Number of central-gland slices	P-value
$Corr_{auc}$	1	0.042
E_{wm}	1	0.044
E_{auc}	1	0.014
C_e	3	0.040

Texture measure	Number of central-gland slices	Significance
E_e	1	<5%
E_{auc}	1	<5%

Table 1: Texture measures that separate the radiologically defined groups. Corr, E and C are Correlation, Energy and Contrast measures of the co-occurrence matrices. The subscripts refer to the parameterisations of the scale-signatures: auc – area under the curve, wm – weighted mean, e – entropy. i.e. E_{auc} is Energy parameterised by area under the curve. The top table is derived from features that are normally distributed. The lower table is derived from features that are not normally distributed. These were evaluated using a Mann-Whitney test and the parameters derived corresponded to tabulated significance values < 5%.

Discussion and Conclusions

The definitive test to discriminate between stromal and non-stromal BPH is histological examination of biopsy samples. Histological data were not available for these images. The radiological classification used a set of published criteria which have been themselves validated against histology. There are many texture measures in the literature, and it would require an extensive project to establish the most appropriate ones for this application. Co-

occurrence matrices have the virtue that they can, in principle, capture a large range of spatial relationships among pixel values and they allow a number of different features to be evaluated within the same framework. Of the features tested, only a few resulted in significant classification performance. These were mainly associated with the Energy measure of the co-occurrence matrix, perhaps indicating that this measure maps well onto the underlying texture associated with glandular BPH. The parameters derived from the scale-signatures to be used as features were rather heuristic, and do not leave much room for interpretation.

However, the features do relate to image texture and its scale-dependence. Importantly they *only* relate to image texture, and not to any other features such as the size of the central gland or its ratio to total prostate volume. Selection of the sampling area to be no larger than the smallest central gland in the image set ensures that only central-gland features are sampled This is an encouraging sign, indicating that there is scope for image measures that can discriminate between different components of BPH. Given a differential segmentation of the prostate, the ratio of central gland to total prostate volume would also be available, which is a key component in the criteria published by Ishida et al. [5].

References

[1] A. M. Alam, K. Sugimura, H. Okizuka, *et al.*, "Comparison of mr imaging and urodynamic findings in benign prostatic hyperplasia", *Radiation Medicine,* **18,** pp. 123-128, 2000.

[2] A. Tewari, K. Shinohara and P. Narayan, "Transition zone volume and transition zone ratio - predictor uroflow response to finasteride therapy in benign prostatic hyperplasia patients", *Urology,* **45,** pp. 258-264, 1995.

[3] P. D. Allen, D. Williamson, J. Graham and C. E. Hutchinson, "Differential segmentation of the prostate in mr images using tissue modelling and 3d active shape models", *IEEE International Symposium on Biomedical Imaging* 2006 (J. Kovacevic and E. Meijering, ed.) ed.) pp 410-413.

[4] K. Isen, Z. Sinik, T. Alkibay, *et al.*, "Magnetic resonance imaging and morphometric histologic analysis of prostate tissue composition in predicting the clinical outcome of terazosin therapy in benign prostatic hyperplasia", *International Journal of Urology,* **8,** pp. 42-48, 2001.

[5] J. Ishida, K. Sugimura, H. Okizuka, *et al.*, "Benign prostatic hyperplasia - value of mr-imaging for determining histologic type", *Radiology,* **190,** pp. 329-331, 1994.

[6] M. L. Schiebler, J. E. Tomaszewski, M. Bezzi, *et al.*, "Prostatic-carcinoma and benign prostatic hyperplasia - correlation of high-resolution mr and histopathologic findings", *Radiology,* **172,** pp. 131-137, 1989.

[7] P. D. Allen, H. L., J. Graham and C. E. Hutchinson, "Identifying tissue types in mr images of the prostate", *Medical Image Understanding and Analysis* 2005 (M. Mirmehdi, ed.) ed.) pp

[8] R. Haralick, "Statistical and structural approaches to texture", *Proceedings of the IEEE,* **67,** pp. 786-801, 1979.

A Comparative Evaluation of Cortical Thickness Measurement Techniques

P. A. Bromiley*, M.L.J. Scott, and N.A. Thacker

Imaging Science and Biomedical Engineering, Stopford Building,
University of Manchester, Oxford Road, Manchester, M13 9PT.

Abstract. *In vivo* measurements of cortical thickness from MR images have potentially widespread utility in the characterisation of normal brain development and maturation as well as in diagnosing and measuring the progress of a number of cortical pathologies. The literature describes several approaches to this problem, which may be divided into two groups: those relying on deformable models of the inner and outer cortical surfaces, and those relying on image intensities alone. Results from the former may be largely model driven at points deep within the sulci, where no apparent channel of cerebrospinal fluid can be seen at the resolution of typical MR images, potentially introducing bias. We present a comparative evaluation of cortical thickness measurement techniques, which demonstrates that approaches based on edge detection can provide cortical thickness measurements as accurate as those from model-based approaches, using less processor time, and without the possibility of bias from a model.

1 Introduction

The human cerebral cortex makes up the largest part of the brain, and consists of a highly convoluted layer of neuronal cells with the topology of a 2D sheet, surrounding a core of white matter. Its thickness varies considerably, from approximately 2mm in the calcarine sulcus to approximately 4mm in the precentral gyrus, with an average of approximately 3mm [1]. Measurements of cortical thickness have shown considerable potential both in the study of normal brain growth and maturation [2–6] and in the diagnosis, or measurement of the progression, of a wide variety of cortical pathologies including Alzheimer's disease [7], schizophrenia [7], and others [6]. Reliable and automated techniques for *in vivo* cortical thickness measurement therefore form a useful tool in neurology.

Any cortical thickness measurement technique requires two core components: a method for locating the inner and outer cortical surfaces, and a metric with which to measure the distance between them. A variety of thickness metrics have been suggested, varying from simply measuring the thickness along normals to the inner cortical surface [8] to approaches based on partial differential equations [9–11]. However, the thickness metric chosen must be considered the definition of the quantity to be measured: the accuracy with which this measurement is made is dictated by the method chosen to define the cortical surfaces. The techniques presented in the literature can be coarsely divided into two groups in this respect: model-based approaches that involve fitting deformable models to the inner and outer cortical surfaces i.e. the boundaries between white matter (WM) and grey matter (GM) and GM and pia matter (PM), and data-driven approaches that detect these interfaces using image intensities alone. The deformable model based approach typically involves segmenting the WM, fitting a model to the WM/GM interface, and then expanding this surface until it reaches the GM/PM interface. Many such algorithms include topological constraints that prevent self-intersections in the models [8, 12], thus ensuring that they have a simple, spherical topology, which has advantages in defining the distance metric and allows the results to be displayed as projections onto the cortical surface. This approach is used in the popular ASP (Anatomic Segmentation using Proximities) algorithm [8], which introduces a surface self-proximity term and a term based on the distance between corresponding vertices on the inner and outer surface models.

The deformable model based approach has two main drawbacks. First, such algorithms require considerable computational resources, largely due to the topological constraints. For example, the algorithms presented in [8] and [12] required 30hrs on a 180 MHz Silicon Graphics R10000 and 5 hours on a 500 MHz Pentium 3 respectively to process each data set. Second, the introduction of terms to prevent self-intersections of the model surfaces may bias the algorithm towards a fixed separation between the inner and outer cortical surface models [8], depending on the weights assigned to each term. Such terms may be required in order to solve the notoriously difficult problem of fitting the model in tightly folded gyri [13, 14], where the sulcal banks oppose so closely that there is no clear CSF channel in the sulcus at the resolution of typical MR images. In such regions the outer surface model may fail to fit the pial surface within the sulcal fundus, leading to thickness estimates that are at least two times too high. Such biases may therefore be inevitable. These problems have led to an interest in data-driven techniques, in which the inner and outer cortical surfaces are determined using only image intensities (e.g. [11, 14, 15]). Such approaches also typically begin with segmentation of the WM, GM and cerebrospinal fluid (CSF). Thickness measurements are then performed at each

* E-mail: paul.bromiley@man.ac.uk

point on the inner cortical surface, by propagating away from the surface according to some thickness metric until the outer cortical surface is reached. One notable advantage of such approaches is that, in regions of tightly folded gyri as described above, if the pial surface is missed and another point on the inner cortical surface is reached, the thickness measurement can be halved to produce a value which is approximate but still entirely data-driven and so free from model bias. It has been shown that this approach has little effect on final, regional thickness measurements if local smoothing is applied [11, 14].

In previous work we have presented a data-driven cortical thickness measurement technique [16]. In this paper, we compare the results from this algorithm to a wide range of published measurements, produced using both model-based and data-driven algorithms. The aim was to determine whether the use of deformable models to define the inner and outer cortical boundaries increases the accuracy of the thickness measurement (i.e. reduces the random errors) or introduces bias (i.e. systematic errors).

2 Method

The cortical thickness measurement technique used in this study has been presented previously [16]; we summarise the method here. Several stages of preprocessing were used. Initially, the partial volume segmentation technique described in [17], which involves fitting a Bayesian mixture model containing both pure tissue and partial volume terms to the image histogram, was applied to the GM, WM and CSF. This process produced measurements of the means and standard deviations of the pure tissue grey levels and measurements of the most likely tissue volume contributions in each voxel. To obtain a finer through-plane resolution whilst preserving tissue boundaries, the data was explicitly up-interpolated using a partial volume scheme to constrain the potential tissue boundaries, determined using 3D image gradients, that could pass through a partial volume voxel. Finally, a map of the GM was produced. In addition, the original image data was registered to a common stereotaxic space (the Talairach atlas [18]) using a linear affine transform. The atlas defined the 31 cortical regions used later in producing regional histograms of the cortical thickness: its use allowed direct comparison to results presented in [1].

The cortical thickness was measured using a modified edge detection process (Canny [19]) to determine the GM/WM boundary. A 'z-score' measure of the grey-level of each voxel being consistent with the GM/WM midpoint value was used to construct a likelihood image which highlighted the GM/WM boundary. This was then used as a replacement for the conventionally used sum-squared image gradient (edge strength) map in the Canny edge detector in order to produce well localised connected edge strings to sub-pixel accuracy. The 3D surface normal at each voxel on the GM/WM boundary was determined by taking the local grey-level gradient of the 3D, Gaussian-smoothed (using a kernel of [1/2, 5]), up-interpolated grey-level data. The GM tissue probability maps and the GM/WM boundary and edge orientations were used to determine the distance from the boundary, at each voxel on the boundary, along the orientation direction to another GM edge. This edge could either be a GM/CSF interface or, if there was no CSF visible in the intervening sulcus, a GM/WM interface, and was determined by comparing the value of the voxel in the original grey-level image to the mean GM value. If it was a WM boundary it was assumed that the opposing banks of a sulcus had been traversed and the sulcal thickness was taken to be half this length: this approach has been shown to have little effect on subsequent regional thickness measurements [11, 14]. Histograms of the cortical thickness measurements in each region were then produced and their median values taken to produce the final regional thickness measurements. The regional histograms typically contained between 100 and 6000 entries, thus giving reasonable anatomical precision as well as a robust estimate of the median.

The method was applied to inversion-recovery MR scans (1.5T Philips ACS PT 6000 NT, TI/TR/TE = 300/6850/18 ms, pixel size = 0.9x0.9mm, 51 slices) of 119 normal volunteers (52 male, mean age = 70.3 years, range = 19-86 years). 110 scans had axial slices (thickness = 3.0mm), 9 had coronal (thickness = 4.0mm). The results were then compared to cortical thickness measurements from the literature in two stages. First, regional average thickness measurements in a range of standard cortical regions were computed and compared to the results published by Kabani et al. [1], which were produced using a deformable model-based technique. Only the youngest 13 subjects (mean age 36.9 years, range 19-53) were included, to ensure that the mean age of the group was similar to that used by Kabani et al. Linear fits to cortical thickness vs. age in all 110 subjects were used to correct for the remaining age difference. This analysis acted as an exemplar comparison between model-based and data-driven techniques. Second, a meta-study of precentral gyrus thickness variation with age was performed, involving measurements from nine studies including this one, and incorporating 635 subjects. This region is the location of the primary motor cortex, and is also the thickest region of the cortex; however, the choice of region was solely based on the number of previously published thickness measurements available in the literature. It is known that changes in both cortical thickness [10] and overall brain volume [20] occur with age. Any bias introduced by the use of deformable models would suppress the ability to observe such changes,

and so this meta-study provides a more detailed evaluation of such biases.

3 Results

Kabani et al. [1] presented regional average cortical thickness measurements in ten regions from 40 subjects produced using the ASP algorithm. Half of the subjects were used in each hemisphere, giving an effective group size of 20. In addition, they presented manual measurements of the same quantities. Figure 1a shows a comparison of these results to thickness measurements in the same regions produced using the algorithm presented here. Assuming that the manual measurements provide a gold standard, Fig. 1b provides a more quantitative comparison, showing the difference between the manual measurements and the results from each algorithm. The mean differences across all regions are 0.61 ± 0.43 mm for the ASP algorithm and -0.21 ± 0.22 mm for the algorithm presented here. Neither difference is statistically significant, and so there is no evidence of systematic error in either algorithm. Given that the group sizes are similar in each case (20 in Kabani et al. vs. 13 in the present study) the statistical errors from the algorithm presented here are approximately half as large as those on the Kabani et al. results.

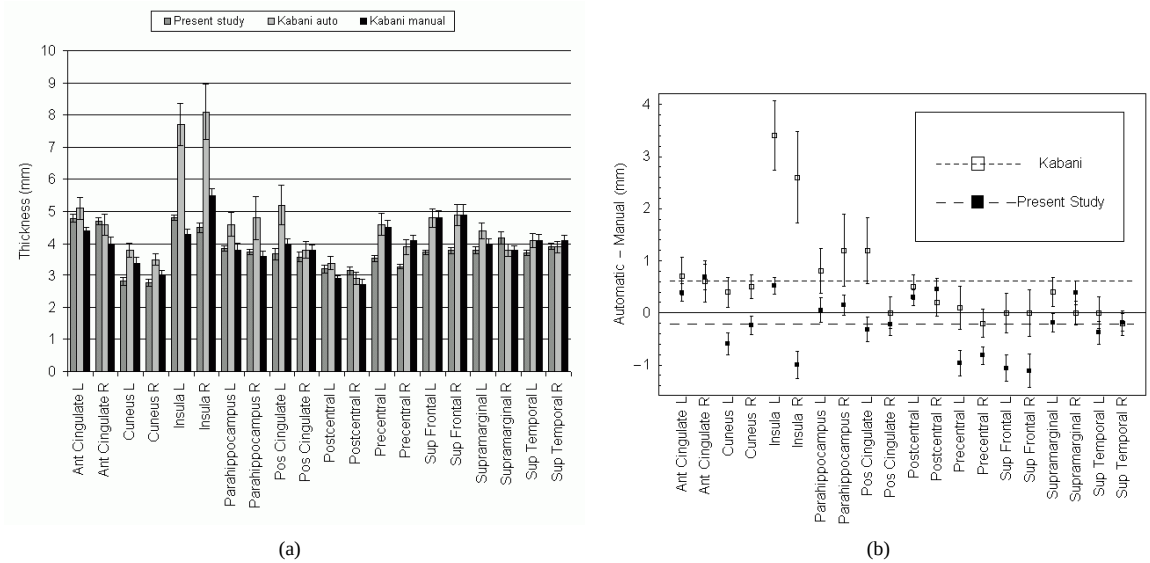
Reference	No. subjects	Age range (years)	Algorithm type
Kabani et al. 2001 [1]	40	18-40	Model based
von Economo 1929 [21]	-	30-40	Manual measurement
Sowell et al. 2004 [10]	45	5-11	Intensity based
Tosun et al. 2004 [22]	105	59-84	Model-based
Fishl et al. 2000 [12]	30	20-37	Model-based
Thompson et al. 2005 [14]	40	18-48	Intensity based
MacDonald et al. 2000 [8]	150	18-40	Model based
Salat et al. 2004 [5]	106	18-93	Model based

Table 1: Details of the studies included in the meta-study of the dependence of precentral gyrus thickness on age.

Figure 2 shows the average cortical thickness measurements for the precentral gyrus in 119 subjects produced using the algorithm presented here, plotted against age. A quadratic fit to the data is shown: the dashed curves either side of the fit show the upper and lower standard error bounds. A significant ($P < 0.0001$) reduction of cortical thickness with age is observed. In previous work [16] we found similar dependencies in other cortical regions. Also shown are a number of measurements of precentral gyrus thickness from the literature. The details of the studies involved are given in Table 1. With the exception of the results from Kabani et al. [1], the data were read from graphical representations[1]. The results presented by von Economo [21] were measured manually post-mortem: brain volume decreases by approximately 10% during postmortem fixation [23]. However, the thickness measurement was performed only on the gyral cap, which is known to be thicker than the sulcal fundi [8]. Similarly the presentation of the results in Sowell et al. [10] and Thompson et al. [14] as projections onto the outer cortical surface prevented identification of the thickness in the sulcal fundi. These three data therefore represent upper limits on the average thickness in the region. Overall, the studies represented in Fig. 2 represent the widest possible range of methods for defining the inner and outer cortical surfaces, the thickness metric, and the presentation of the results. Some variation between the measurements introduced by these differences in experimental procedure might therefore be expected. However, the data show a remarkable level of agreement. If the errors on the results quoted by Kabani et al. [1] can be taken as representative of the errors on the other studies, then there is no statistically significant difference between any of these results and our own.

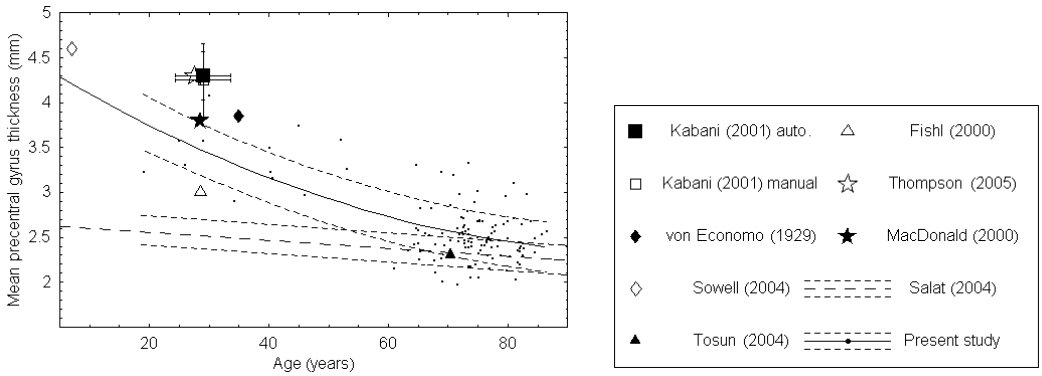
The remaining paper included in this study, Salat et al. [5], was the only one to study variation in cortical thickness with age, and the only model-based study to cover the whole age range between adolescence and senescence. A significant disagreement between this study and the others included in the meta-study can be seen, with Salat et al. suggesting a much lower rate of cortical thickness change with age. Given the level of agreement between the other studies, and the fact that the algorithm used by Salat et al. was based on deformable modelling of the inner and outer cortical surfaces, we suggest that this may be due to the possibility for bias identified by [8] in such algorithms. The constraints applied to prevent self-intersection of the models and to aid modelling of the surfaces in tightly folded gyri tend to bias the model towards a fixed thickness, suppressing the observation of age-related changes. The same effect was found in the mean rate of thickness change over the whole cortex: a global rate of 0.016 ± 0.0052 mm year^{-1} was found in the present study, which is consistent with the $0.0077\ mm\ year^{-1}$ found in [2], but is an order of magnitude higher than,

[1]With the exception the results from Salat et al. [5] and von Economo [21] these were presented as views of the outer cortical surface, in some cases partially inflated to reveal the sulcal fundi, with colour coding to represent the thickness at each point. This method of data display is popular in the literature as it avoids the need for parcellation of the data into particular regions. However, the calculation of regional average thicknesses from such representations is difficult and the calculation of errors on the averages impossible. Hence, these points are shown without error bars.



(a) (b)

Figure 1. Manual and automatic regional average cortical thickness measurements presented in [1], and automatic results from the algorithm presented here (a), and differences between the manual thickness measurements and the algorithm results (b); the dashed lines show average differences across all regions for each algorithm.



(a) (b)

Figure 2. Measurements of the average cortical thickness in the precentral gyrus: see main text for description.

and inconsistent with, the 0.0016 mm year^{-1} rate quoted by Salat et al. Finally, age-related changes in GM density i.e. the proportion of GM within a kernel around each point on the cortical surface, a quantity closely related to cortical thickness, were measured in [4]: the average, proportional rate of change of GM density in the precentral gyrus was $0.593\ \%\ year^{-1}$, consistent with the $0.567 \pm 0.270\ \%\ year^{-1}$ cortical thickness change found in the present study.

4 Conclusion

The literature describes several approaches to cortical thickness measurement, many of which rely deformable models fitted to the inner and outer cortical surfaces. The topological constraints applied in such algorithms require considerable amounts of processor time, typically several hours per image volume. In addition, bias may be introduced by the model at points where the information available from the image is weak. In previous work we presented a simpler alternative, based on edge detection followed by measurement of the thickness along normals to the inner cortical surface, which requires an order of magnitude less processor time and does not introduce bias. We have presented comparisons of the results from this algorithm against a range of measurements from the literature. The comparisons indicate that results from this algorithm are more accurate than those published in the literature, suggesting that the accuracy of both types of algorithm is dictated by the initial segmentation. However, in at least one study, the use of an algorithm based on deformable models appears to have suppressed the ability to detect age-related change. We therefore conclude that the use of deformable models in cortical thickness measurement provides no advantage in terms of accuracy.

The comparison of age-related changes in cortical thickness can also be used to investigate the ability of the algorithm presented here to detect pathological effects. Comparison with previous results shows that the thickness measurements in the precentral gyrus are consistent with those presented in the literature over a wide age range, confirming the accuracy of the technique. In previous work [16] we have shown that statistically significant age-related changes were detected in this data set throughout the cortex. Previous studies on age-related and pathological brain volume change [24] have shown that pathological changes typically occur an order of magnitude faster than age-related changes. Therefore, the ability to detect age-related changes strongly implies that pathological changes will also be detected.

Acknowledgements

This work was supported by the University of Manchester Translational Imaging Unit and funded by the MIAS (Medical Images and Signals) IRC under EPSRC grant no. GR/N14248/01 and the UK Medical Research Council grant no. D2025/31, and by the Wellcome Trust grant no. ME003889. The software used in this study is freely available from www.tina-vision.net.

References

1. N. Kabani, G. Le Goualher, D. MacDonald et al. "Measurement of cortical thickness using an automated 3-D algorithm: A validation study." *NeuroImage* **13**, pp. 375–380, 2001.
2. V. A. Magnotta, N. C. Andreasen, S. K. Schultz et al. "Quantitative in vivo measurement of gyrification in the human brain: Change associated with aging." *Cerebral Cortex* **9**, pp. 151–160, 1999.
3. E. R. Sowell, P. M. Thompson, K. D. Tessner et al. "Mapping continued brain growth and grey matter density reduction in dorsal frontal cortex: Inverse relationships during postadolescent brain maturation." *J. Neurosci* **21**, pp. 8819–8829, 2001.
4. E. R. Sowell, B. S. Peterson, P. M. Thompson et al. "Mapping cortical change across the human life span." *Nature Neuroscience* **6**, pp. 309–315, 2003.
5. D. H. Salat, R. L. Buckner, A. Z. Snyder et al. "Thinning of the cerebral cortex in aging." *Cerebral Cortex* **14**, pp. 721–730, 2004.
6. A. W. Toga, P. M. Thompson & E. R. Sowell. "Mapping brain maturation." *Trends in Neurosciences* p. In Press, 2006.
7. P. M. Thompson, K. M. Hayashi, E. R. Sowell et al. "Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia." *NeuroImage* **23**, pp. S2–S18, 2004.
8. D. MacDonald, N. Kabani, D. Avis et al. "Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI." *NeuroImage* **12**, pp. 340–356, 2000.
9. S. E. Jones, B. R. Buchbinder & I. Aharon. "Three-dimensional mapping of cortical thickness using Laplace's equation." *Human Brain Mapping* **11**, pp. 12–32, 2000.
10. E. R. Sowell, P. M. Thompson, C. M. Leonard et al. "Longitudinal mapping of cortical thickness and brain growth in normal children." *J. Neurosci.* **24**, pp. 8223–8231, 2004.
11. E. Luders, K. L. Narr, P. M. Thompson et al. "Hemispheric asymmetries in cortical thickness." *Cerebral Cortex* p. In Press, 2006.
12. B. Fischl & A. M. Dale. "Measuring the thickness of the human cerebral cortex from magnetic resonance images." *PNAS* **97**, pp. 11050–11055, 2000.
13. J. S. Kim, V. Singh, J. K. Lee et al. "Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification." *NeuroImage* **27**, pp. 210–221, 2005.
14. P. M. Thompson, A. D. Lee, R. A. Dutton et al. "Abnormal cortical complexity and thickness profiles mapped in Williams syndrome." *J. Neurosci.* **25**, pp. 4146–4158, 2005.
15. G. Lohmann, C. Preul & M. Hund-Georgiadis. "Morphology-based cortical thickness estimation." In C. J. Taylor & J. A. Noble (editors), *Information Processing in Medical Imaging, 18th International Conference, IPMI 2003, Ambleside, UK, July 20-25, 2003, Proceedings*, volume 2732 of *Lecture Notes in Computer Science*, pp. 89–100. Springer, 2003.
16. M. L. J. Scott & N. A. Thacker. "Regional cerebral cortical thickness measurement using MRI." In *Proc. MIUA*, pp. 7–10. 2005.
17. M. Pokrić, N. Thacker, M. L. J. Scott et al. "Multi-dimensional medical image segmentation with partial voluming." In *Proc. MIUA*, pp. 77–81. 2001.
18. J. Talairach & P. Tournoux. *Co-planner stereotaxic atlas of the human brain. 3-Dimensional proportional system: An approach to cerebral imaging*. Thieme Medical Publishers, New York, 1988.
19. J. F. Canny. "A computational approach to edge detection." *IEEE PAMI* **8(6)**, pp. 679–698, Nov 1986.
20. P. A. Bromiley, N. A. Thacker & A. Jackson. "Trends in brain volume change with normal ageing.", 2005. Proc. MIUA.
21. C. V. Economo. *The Cytoarchitectonics of the Human Cerebral Cortex*. Trasnlasted by S. Parker. Oxford University Press, Oxford, 1929.
22. D. Tosun, M. E. Rettmann, X. Han et al. "Cortical surface segmentation and mapping." *NeuroImage* **23**, pp. S108–S118, 2004.
23. R. Quester & R. Schroder. "The shrinkage of the human brain stem during formalin fixation and embedding in paraffin." *J. Neurosci. Methods* **75**, pp. 81–89, 1997.
24. N. C. Fox, S. Cousens, R. Scahill et al. "Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer's disease." *Archives of Neurology* **57**, pp. 339–344, 2000.

Groupwise registration of richly labelled images

K O Babalola* and T F Cootes

Division of Imaging Science and Biomedical Engineering (ISBE), University of Manchester, UK

Abstract.
We describe a method of registering 3D images in which regions have been segmented and labelled. Standard registration schemes cannot be naively applied to such images and several modifications to allow their registration have been proposed such as using vector valued images with one plane for each label [1] or a *label consistency* measure [2]. However, these either lead to impractically large images when a large number of labels are being considered, or cannot be applied to groupwise registration in a straightforward manner.
The method we describe does not lead to impractically large images and can be applied to both groupwise and pairwise registration. It involves mapping each label value to a vector in a low dimensional space and applying a multi-plane registration algorithm to the resulting vector valued image. To obtain good results, the vectors representing each label should be well separated and chosen in such a way that there is minimal confusion between them. We demonstrate the method by using it to register a set of richly labelled images in a groupwise manner and use the resulting correspondences to construct a statistical shape model of a number of subcortical structures in the brain.

1 Introduction

Non-rigid registration is widely used to find correspondences between images, and between atlases and images [3, 4]. Many techniques have been proposed for registering unlabelled grey level images. Where the images have been segmented, so that each voxel contains a label indicating which structure it belongs to, one commonly has a single plane image of label values (such as an integer indicating the region type). Registering labelled images is a necessary step in the automatic construction of statistical shape models. However, standard registration algorithms should not be naively applied, as the difference between voxel values (integer labels) is no longer a meaningful metric. Instead, Tsai *et al.* [1] suggested extracting binary planes, one for each structure and encode the structures in a multi-dimensional binary vector-valued image. Thus if each voxel can be labelled with an integer in the range $[1, N]$, then N planes are created.

We propose an alternative technique. We generate a mapping $i \rightarrow \vec{v}_i$, which replaces each label index i with a vector value $\vec{v}_i$ in an m-dimensional space. The dimensionality, m, can be any positive integer, but in practise will be kept relatively small (see below). The key to the approach is to ensure that the vectors $\{\vec{v}_i\}$ are well separated in the space so that there is minimal interference between labels from different structures when they are nearby in the image. Having applied the mapping, we can smooth the images if required to encourage convergence, and then apply any multi-plane registration algorithm.

In the following we will describe the approach, showing one choice for the mapping, and use it to register a set of richly labelled images with a "groupwise" algorithm [5], which seeks to find the best registration of a group, rather than just a pair of images. Section 2 gives details of related work and section 3 gives details of the mapping process. The construction of SSMs using the modified group-wise method of Cootes *et al.* [5] is described in section 4 and quantitative analysis of the results of the registration are shown in section 5. The paper ends with a discussion.

2 Related work

Frangi *et al.* [6] propose a method of automatically constructing 3D Active Shape Models from binary images of a particular structure. This required the registration of binary images of the structure being modelled. They used a *quasi-affine* registration with nine degrees of freedom and a normalised mutual information metric to perform an initial registration to an atlas, then combined the shapes using their signed distance transforms. This method gave promising results, however, its application to the case of images containing more than one structure was not investigated.

Frangi *et al.* [2] construct statistical shape models of the ventricles of the heart from labelled images. They define two similarity measures applicable to registration of a pair of labelled images. Both vary between 0 (the worst value) and 1 (the best value). The first, *Label consistency*, measures the fraction of labels in the source shape that are correctly mapped into the target shape. The second, the κ *statistic*, is used in biomedical research to assess inter-rater agreement.

*Email: kola.babalola@manchester.ac.uk

These measures allowed the statistical shape models to be constructed, however, they are only defined for pairwise registration, and their extension to groupwise registration has not been explored.

Tsai *et al.* [1] built shape models from images with labels of multiple structures and applied the models to segmentation. However, they constructed their models by creating binary vector-valued images, with the dimensions of the vectors determined by the number of unique labels. They registered the images using an overlap based cost function and constructed the shape models using signed distance transforms of the aligned binary shapes. This approach results in memory limitations restricting the number of labels that can be considered during registration.

3 METHODS

3.1 Mapping labels to vectors

We seek to map each label value i to a unique vector, $\vec{v}_i$ in an m-dimensional space, so as to ensure that different resulting vectors are well separated. However, we also wish to arrange that if the resulting vector image is smoothed (either explicitly in order to improve registration convergence, or implicitly when interpolating values), there is a minimal chance of vector values being created which could be confused with other labels. Consider the border between two structures, labelled i and j. On one side of the border voxels are mapped to vector $\vec{v}_i$, and on the other to $\vec{v}_j$. If the vector image were smoothed, voxels around the border would have values lying on the line segment between $\vec{v}_i$ and $\vec{v}_j$, i.e. $\vec{v} = \alpha\vec{v}_i + (1-\alpha)\vec{v}_j$. We would thus like to ensure that no other labels have values near that line. Similarly at junctions between three structures, blurring produces vectors which are linear combinations of three key vectors, in a triangular region.

Ideally we would thus like to select vector values $\{\vec{v}_i\}$ which are both well separated, and such that the linear combinations of pairs or triplets which occur at boundaries are unlikely to be mistaken for any other values.

A complete solution would involve analysing the original images to determine which structures are neighbours, and using this as a constraint in an optimisation. However, as an initial approximation we have experimented with a mapping in which the background label is mapped to zero ($\vec{v}_0 = \vec{0}$), and all other labels are mapped to points maximally separated on a unit hypersphere ($|\vec{v}_i| = 1$).

We have treated the background differently from other labels because in the general case any label could be a neighbour of the background. Without making use of neighbour information, this arrangement should lead to the least interference between different structures. However, there is a possibility of weighted sums of non-background labels representing the background. This can be averted by taking information of label neighbours into account.

3.1.1 Maximally separable points on a hypersphere

Mapping labels to 2D values is equivalent to equally spacing the labels on the radius of a unit circle. For 3D it is equivalent to equally spacing the labels on the surface of a unit sphere. For the 2D case calculating the positions are trivial. For the 3D case one could use the analogy of equally spaced charges.

However, the general case of spacing m points equally on the surface of a unit hypersphere is equivalent to the optimisation problem :

$$\text{Find } m \text{ points} \quad \{\mathbf{x}_i\},\ |\mathbf{x}| = 1,\ i = 1 \ldots m \quad \text{maximally separated} \tag{1}$$

$\mathbf{x}_i.\mathbf{x}_j = \cos\theta$ as $|\mathbf{x}_i| = |\mathbf{x}_j| = 1$ and $\mathbf{x}_i.\mathbf{x}_j = 1$ if $\mathbf{x}_i = \mathbf{x}_j$ and $\mathbf{x}_i.\mathbf{x}_j < 1$ if $\mathbf{x}_i$ is away from $\mathbf{x}_j$. We therefore seek to minimise an objective function of the form :

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbf{x}_i.\mathbf{x}_j \qquad \text{subject to } |\mathbf{x}_i| = 1 \quad \forall\, i \tag{2}$$

Performing such an optimisation is not trivial due to the sinusiodal nature of the *dot* product. At the moment we use an iterative procedure to equally space our points. We generate the vectors of the appropriate dimensionality for the

required number of points from a random multidimensional distribution and normalise them to unit length. We then iteratively find the two closest points in the set and increase their separation by a small distance normalising both vectors afterwards. This is repeated until the standard deviation of the closest distance between the points falls below a given value, or a maximum number of iterations is reached. Figure 3.1.1 shows the results of mapping 20 points to 2D and 3D using this algorithm.

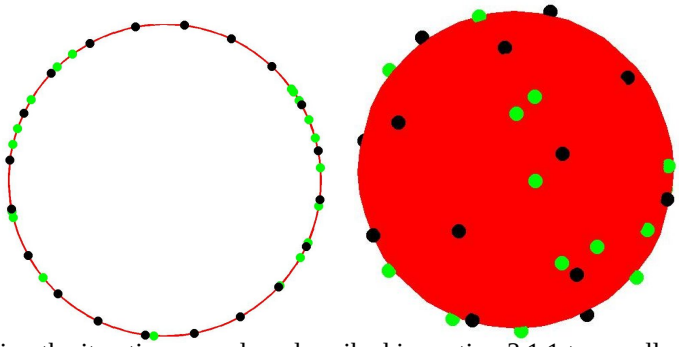


Figure 1. Results of applying the iterative procedure described in section 3.1.1 to equally spacing 20 points on a unit circle in 2D (left) and on a unit sphere in 3D (right). The green circles show the initial distribution of the random points, and the black circles show the final result.

3.2 Group-wise registration

It is often necessary to construct statistical models of the shape and shape variation of structures appearing in medical images. This requires finding correspondences across a set of images. Rather than use pair-wise techniques (such as selecting one image as a reference and independently registering each of the other images to it), there is considerable interest in *groupwise* methods, which attempt to find the optimal correspondences across the group as a whole [5, 7, 8]

We adopt a variant of the minimum-description length approach [5, 8] in which the quality of the current registration is evaluated by estimating the amount of information that would be required to encode the training images using a statistical model of shape and texture constructed using the current correspondences. This gives an objective function (a description length), which can be optimised by manipulating the correspondences.

Correspondence is represented by the position of a set of control points on each image (piece-wise linear interpolation is used to estimate correspondence away from the control points). The general approach is as follows (though see [5] for details):

1. Select one image to be used as an initial reference
2. Initial affine registration to reference:
   (a) For each image in turn estimate (affine) movement of control points to optimise match to reference image
3. Groupwise registration:
   (a) For each image in turn
      i. Construct a shape and ‘texture’ model using current points in all other images
      ii. Modify the positions of the points in the target image to minimise the cost of encoding that image using the current model
   (b) Repeat until convergence

In this case ‘texture’ refers to vector values at each voxel.

4 EXPERIMENTS

We applied group-wise registration as described above to register 37 subjects from a dataset provided by the Centre for Morphometric Analysis, Boston. This consisted of T1 MR images and corresponding labelled images of each subject. 35 cortical and subcortical structures in the images had been labelled. We mapped label values to 3 dimensions and registered the 3 plane images of each subject. The transformations obtained from these registrations were used to establish correspondence between the surfaces of 10 subcortical structures (accumbens, amygdala, brain stem, fourth ventricle, caudate, hippocampus, lateral ventricles, pallidum, putamen and thalamus). These formed a training set from which a statistical shape model (SSM) was constructed as described in [9].

5 RESULTS

To quantitatively assess the quality of the registration a distance metric giving the registration error for each subject was obtained as follows. One subject was selected as a template and surfaces for each of the structures listed in section 4 were obtained. The surface of each structure was warped to approximate the surface of the structure in each member of the training set using the correspondences obtained during the registration. The Euclidean distance between the closest points on the original surface to each point on the warped surface was taken as the registration error of the point. If the registration was perfect these distances would be expected to be zero. Figure 2 shows a histogram of the registration error for each of the 37 subjects. A SSM is a linear model approximating a class of shapes by a main shapes and the main ways in which the shapes vary (modes of variation). The first three modes of variation of the resulting statistical shape model are shown in figure 3.

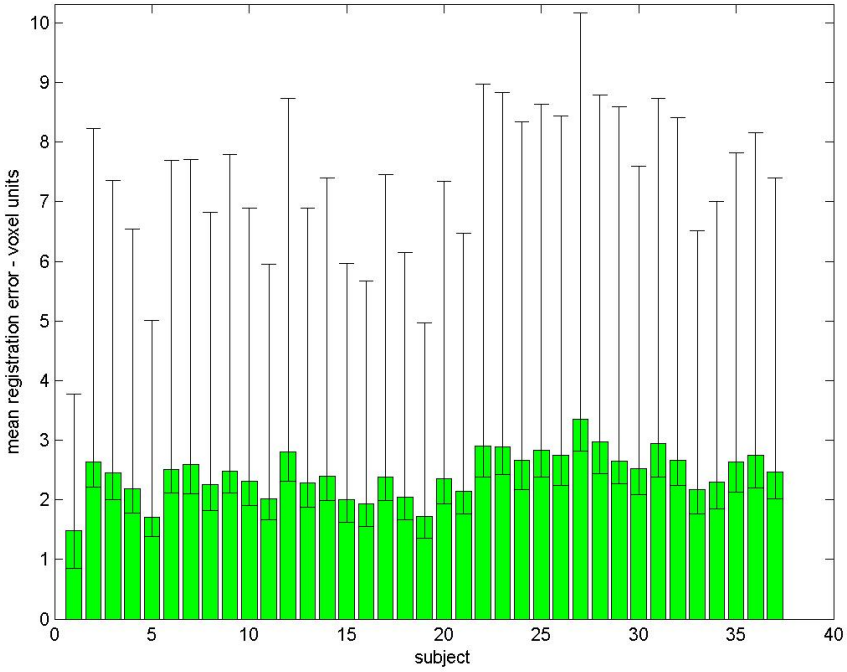


Figure 2. Mean registration error for each subject in the image set. The bars show the mean of the Euclidean distance between densely sampled points on the registered surface and the original surface. The vertical limits show the 10^{th} and 90^{th} percentiles of the point to surface error. The distances were measured in voxel units, and the voxel size being 1mm $\times$ 1.5mm $\times$ 1mm

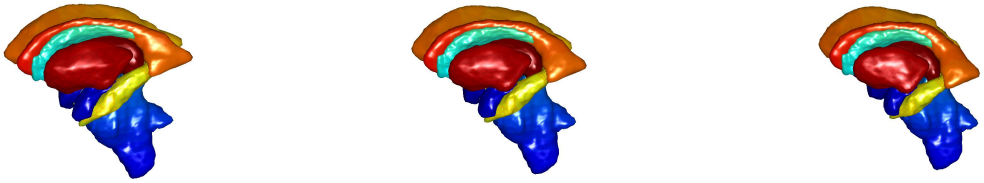
6 DISCUSSION

We have described a method of registering richly labelled data by replacing each voxel label with a vector in a low dimensional space. The values of the vectors should be chosen to ensure that those belonging to different labels are well separated, and that when smoothed there is minimal chance of confusion.

Our initial implementation simply maps points to (randomly assigned) maximally separated vectors on the unit hypersphere. When we use the technique as pre-processing for a groupwise registration algorithm, we obtained encouraging results. In future work we will explore the effect of the choice of dimension m for the vector space, and investigate methods of taking known neighbour relationships into account when assigning vectors to structure labels. We believe inclusion of the neighbour relationships will improve the accuracy of registration. We will also perform quantitative evaluation of the method and compare it with other registration methods, in particular with the pairwise method based on *label consistency* [2].

7 ACKNOWLEDGEMENTS

This work was funded by the EPSRC under the IBIM project, and David Kennedy provided the MR images used.



(a) +3 and -3 standard deviations from mean (centre subfigure) along mode 1

(b) +3 and -3 standard deviations from mean (centre subfigure) along mode 2

(c) +3 and -3 standard deviations from mean (centre subfigure) along mode 3

Figure 3. Mean shape and extremes of variation for the first three modes. The left subfigures are for $-2.5\sqrt{\lambda_i}$ and the right $+2.5\sqrt{\lambda_i}$. λ_i is the variance associated with mode i

References

1. A Tsai, W Wells, C Tempany, E Grimson, and A Willsky, "Mutual information in coupled multi-shape modle for medical image segmentation," *Medical Image Analysis*, vol. 8, no. 4, pp. 429–445.
2. A F Frangi, D Rueckert, J A Schnabel, and Niessen W J, "Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling," *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1151–1166, 2002.
3. K K Bhatia, J V Hajnal, B K Puri, A D Edwards, and D Rueckert, "Consistent groupwise non-rigid registration for atlas construction," in *IEEE Symposium on Biomedical Imaging (ISBI) Arlington*, 2004, pp. 908–911.
4. D Rueckert, A F Frangi, and J A Schnabel, "Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 1014, 2003.
5. T F Cootes, C J Twining, V Petrovic, R Schestowitz, and C J Taylor, "Groupwise construction of appearance models using piece-wise affine deformations," in *Proceedings of 16th British Machine Vision Conference, Oxford*, 2005, pp. 879–888.
6. A F Frangi, D Rueckert, J A Schnabel, and W J Niessen, "Automatic 3D ASM construction via atlas-based landmarking and volumetric elastic registration," *LNCS (IPMI 2001 Proceedings)*, vol. 2082, pp. 78–91, 2001.
7. B Davis, P Lorenzen, and S Joshi, "Large deformation minimum mean squared error template estimation for computational anatomy," in *Proceedings of ISBI*, 2004, pp. 173–176.
8. C J Twining, T F Cootes, S Marsland, V Petrovic, R Schestowitz, and C J Taylor, "A unified information-theoretic approach to groupwise non-rigid registration and model building," in *LNCS (IPMI 2005 Proceedings)*, 2005, vol. 3565, pp. 1–14.
9. T F Cootes, C J Taylor, D H Cooper, and J Graham, "Active Shape Models - their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

Fully automated level set shape prior applied to the foetal heart – preliminary results

Irving Dindoyal[a*], Tryphon Lambrou[a], Jing Deng[a,b], Andrew Todd-Pokropek[a]

Departments of [a]Medical Physics, [b]Obstetrics and Gynecology, University College London, UK

According to our knowledge this paper presents the first shape based constraints applied to foetal echocardiography. In our work we describe a level set model that uses region based information in the form of the Mumford-Shah energy functional and a shape prior term. The prior is a template level set that is constructed as the mean shape from only three datasets. After global registration of the template to unseen images the template constrains the level set to remain within its respective chamber and results in fewer tunnelling through weak or partially imaged walls than in the absence of shape prior constraints. In the presented images the average root mean square errors of the perpendicular distances between the algorithm's delineation and manual tracings are to within 3 pixels with shape prior and under 7 pixels using the level set without shape constraint. Recovery of intra cardiac boundaries with heavily corrupted data is also presented. Though our results are preliminary we expect better performance from additional data incorporated into the prior.

1 Introduction

Congenital heart disease affects about 8 in every 1000 births [1] and its signs can be diagnosed with prenatal echocardiography [2]. As with the adult heart, functional volume estimation of the left ventricle provides quantitative information about the state of the myocardium. However, in the foetus the blood flow in both sides of the heart is allowed to mix and so both ventricles are important for clinical assessment. One important application of foetal cardiac segmentation is for measurement of the absolute chamber size. The prenatal heart has very thin chamber boundaries particularly in the areas consisting of the atrial septum, the membranous segment of the ventricular septum, and the valvular leaflets. Often the resolution of the ultrasound beam perpendicular to its axis is insufficient to resolve these structures and so these walls suffer from signal dropout and appear as holes in the endocardium. These dropouts can also be misleading for clinical diagnosis since the foetal heart contains septal holes which normally close at birth. Artefacts such as these complicate the automated functional volume quantification of each chamber for determining useful cardiac indices such as ejection fraction. In some cases it is difficult for foetal cardiologists to manually trace the endocardiac structures because of the missing image greyscale information.

The Navaux group are the only ones who have considered the use of priors in an algorithm to automatically segment the foetal heart. Their work consists of neural networks [3], [4] whose weights are trained on several datasets and variations of them are used to classify the foetal heart.

This paper presents an implicit level set model with shape prior term motivated by the work in [5] and combines the Mumford-Shah model from [6] to minimise the region-based energy function for image segmentation. Although the work presented here is only for 2D images we expect similar results from application to 3D datasets since level set models can be defined in arbitrary number of dimensions without changing the parameterisation.

2 Method

Segmentation of medical data is prone to errors due to spurious artefacts of the imaging process. Shadowing, attenuation, speckle and partial volume effects are some of the main effects of imaging with ultrasound scanners. Modelling of the images can be used to deconvolve the artefacts and reconstruct the image however; this is a non-trivial task since scans available to the research group are obtained after undisclosed ultrasound scanner reconstruction and non consistent preprocessing have taken place during imaging.

Prior knowledge of the shape of the organ allows for greater tolerance for image artefacts over algorithms that rely on following salient image features such as amorphous deformable models. A template of the organ shape is an approximation of a sample of the population and can be used to guide and constrain the segmentation in areas of image noise where models without prior shape knowledge will fail.

Implicit segmentation by level set models involves solving energy based minimisation of a curve embedded as the zero level set of a higher dimensional surface. Level set methods are notorious for their advantages in topological adaptability and handling self intersection over explicit deformable models. The advantages of using explicit models allow pointwise correspondence to be made when constructing a set of shape outlines. The level set model proposed by Leventon [5] treats the native level set Signed Distance Map (SDM) domain as a component of the statistical

model. Provided all shapes in this domain are placed at the same resolution in SDM space a comparable set of organs can be constructed. This avoids the laborious task of defining explicit point correspondences on the training data in repeatable locations and also allows for topological representations to be statistically modelled through an implicit framework.

2.1 Construction of the shape prior

The image from which the template was created consisted of binary filled tracings of foetal hearts in long axial four chamber view. Overall 26 2D images were taken across a mixture of cardiac phases from 3 datasets. The resampled pixel size varied from 0.3 to 1.46mm across the collection of images. As we are able to collect and manually trace more datasets the template becomes more representative of the population. The images were selected such that all four chambers were visible and appeared to have roughly equal areas to aid the registration algorithm. To align the images we use a multi-resolution method similar to Tsai et al [7] where the binary images were registered using the following cost function in an affine sense

$$E_{align} = \frac{\int_{\Omega} (I_T - I_F)^2 \, d\Omega}{\int_{\Omega} (I_T + I_F)^2 \, d\Omega} \tag{1}$$

where Ω is the image domain, T and F denote the transformed and fixed images respectively. The denominator penalises the transformed image from shrinking to zero. Differential evolution was used as the global optimisation method [8] unlike in Tsai [7] whose registration method is based on local gradient descent.

Multi-resolution rigid registration to the first image in the set was performed on all four cardiac chambers simultaneously to maintain their relative positions in the image. This step may be modified in future to allow individual chambers to be aligned as a refinement of the registration process. The results of the registration can be seen in Figure 1. Since we are not trying to find local corresponding points but to align the shapes globally, rigid registration was sufficient for this task. It may be more beneficial to register in the level set domain since signed double precision values would be used instead of 8 bit unsigned greyscale but we have not tested this.

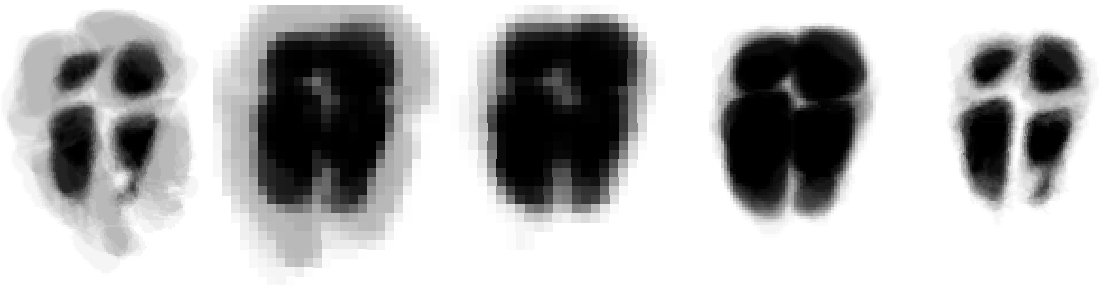


Figure 1 Stages through the registration process illustrated by superposition of manually segmented images. The *first* image shows preregistration. The *second* and *third* images show the intermediate registration at low resolution, the *fourth* at intermediate resolution. The final registered image is shown on the far right at the native resolution. Atria are shown above ventricles in this orientation with the left side of the image corresponding to the left part of the heart and similarly for right hand side.

Each of the registered binary images was transformed into a SDM of the edges u_i of magnitude 1 with negative values inside the shape and positive elsewhere. From each of the SDMs the mean was computed using equation (2).

$$\mu = \frac{1}{n} \sum_{i=1}^{n} u_i \tag{2}$$

2.2 Application of the shape prior to the image

The mean SDM was transformed into a binary image and registered (using the multi-resolution approach above) to the ultrasound image to be segmented using differential evolution and the cost function in equation (1). Only rigid

registration was performed because non-rigid shape analysis would be handled by the image forces in the snake equation later.

The template is automatically fitted to the data after it is placed roughly on the heart. Once the registration is over the snake takes control of the segmentation using the shape template as a starting point. This prevents wasted iterations around seed points and enables faster convergence. In this early implementation the registered template is not free to move after fitting therefore the snakes place a lot of faith in the quality of the registration and the accuracy of the prior. However, the constraints of the template can be controlled via the shape prior weighting coefficients.

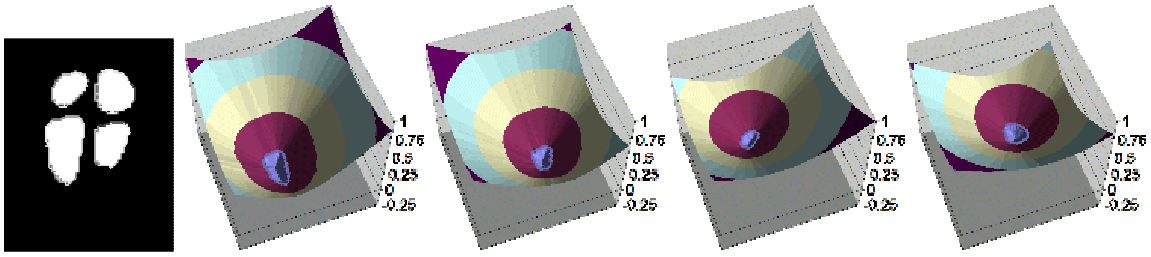


Figure 2 Mean image (shown as binary) and individual chamber shape prior SDMs for each snake (displayed as surface renderings). The surface renderings are shown in left to right order as: left and right ventricles then left and right atria. This figure is best viewed in colour.

The mean image of each chamber was used as the shape prior for the level set snake below. The SDT of each binary image was computed and stored in separated memory spaces as shown in Figure 2.

The multi-chamber snake was constructed from [6] with curvature replacing Sarti's term and a shape prior was added. The snake equation is shown in equation (3) and its evolution is based on mean curvature flow, image region based and shape prior information.

$$\begin{aligned}\dot{\varphi} &= \alpha\nabla\cdot\left[\frac{\nabla\varphi}{|\varphi|}\right]\|\nabla\varphi\| \\ &+\lambda\left(\left[I-\mu_i\right]^2-\left[I-\mu_o\right]^2\right)\exp\left(-\kappa\nabla\cdot\left[\frac{\nabla\varphi}{|\varphi|}\right]\right)\|\nabla\varphi\| \\ &+\beta\left\{\varphi_{sp}-\varphi\right\}\|\nabla\varphi\|\end{aligned} \tag{3}$$

In equation (3) φ is the level set function and the first term is mean curvature flow weighted by α. The second term is taken from [6] and includes the Mumford-Shah (MS) energy functional where I is the current voxel intensity under investigation. μ_i and μ_o are the means of the inside and outside regions of the dataset defined by the level set front.

The MS terms provide the expansion or contraction towards boundaries in the image and use region information dependent on local tissue type. The MS influence models the foreground and background for the image and tries to minimise its energy by separating these two regions. The third term weighted by β in curly brackets is the shape prior contribution. This term returns the signed difference between the two spatially aligned level set functions of the snake φ and shape prior φ_{sp}.

The coefficient β adjusts the relative importance of the shape prior within the snake. To ensure that the shape prior corresponds to the image features the mean shape was registered by differential evolution before the snake was allowed to evolve.

Results

An example of the segmentation both with and without shape prior term enabled is shown in Figure 3. It can be seen that there is less leakage through the thin intra cavity walls with the prior term enabled however, the chamber shapes appear to be slightly rounded to conform to the template. The root mean square error (averaged over all chambers) of the automated contours to manual delineation was computed to quantify the quality of the segmentation.

Segmentation using the shape prior yields noticeably lower average rms errors than in the absence of any shape constraints. Repeatable manual segmentation was considered to have an average rms error of better than 3 pixels although this will vary on the clarity of the dataset however; this assertion has not been tested in our images.

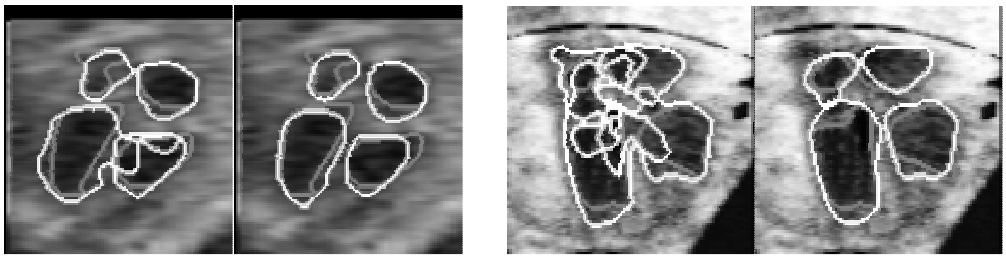


Figure 3 Segmentation after initialisation via registration of the mean template to unseen greyscale images. In these two datasets segmentation without the prior is shown in the left hand image in both cases. Similarly the second image in both datasets was segmented using the shape prior term. *White* contours show automatic segmentation and *grey* contours represent manual tracings. The average rms error of the first dataset is 5.0 and 2.2 pixels for the *left* and *right* images respectively. In the second dataset the average rms errors are 6.2 and 2.5 pixels for without and with shape prior term respectively.

The image in Figure 3 was corrupted with 90% random noise and the template was registered to it. The results of the segmentation both with and without the shape prior term can be seen in Figure 4. When the prior is enabled the snake remains faithful to the template and retains most of the corrupted thin septal walls.

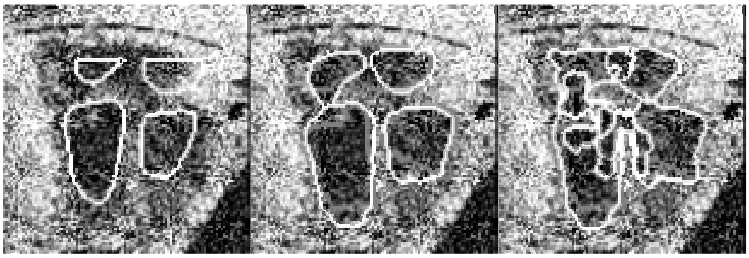


Figure 4 Segmentation of foetal heart image with 90% random noise added. First image shows initial position of the template before registration, segmentation by the snake with shape prior in the *second* image and the *third* shows delineation using the snake only. The automatic segmentations appear in *white* and manual delineation in *grey*. Average rms errors are 3.0 pixels in the *second* image and 5.2 in the *third*.

Conclusion

This paper presents preliminary findings from inclusion of a shape prior term into a level set snake equation. To our knowledge this is the first shape based constraints applied into segmentation of the foetal heart.

The methodology presented in this paper is preliminary and is expected to evolve. We also intend to carry on this work in 3D as more data is collected. The lack of several training examples is restrictive to the analysis presented here such as the ability to distinguish abnormal hearts. We expect the model to perform better using a larger collection of images from which the template is constructed. Validation of the delineation is a step that needs to be addressed in future. We also plan to refine the registration from the evolving segmentation to better fit the template to the image and so provide a more valid prior for the snake.

Acknowledgements

This work was supported by EPSRC (GR/N14248/01) and MRC (D2025/31) under the Interdisciplinary Research Consortium scheme - “From Medical Images and Signals to Clinical Information” (MIAS IRC). Dr Jing Deng is supported by MRC (G108/516).

References

1. Mitchell, S. C., Korones, S. B., and Berendes, H. W., "Congenital heart disease in 56,109 births. Incidence and natural history," *Circulation*, vol. 43, no. 3, pp. 323-332, Mar.1971.
2. Copel, J. A., Gianluigi, P., Green, J., Hobbins, J. C., and Kleinman, C. S., "Fetal echocardiographic screening for congenital heart disease: The importance of the four-chamber view," *British Journal of Obstetrics and Gynaecology*, vol. 157, no. 3, pp. 648-655, 1987.
3. Piccoli, L., Dahmer, A., Scharcanski, J., and Navaux, P. O. A. Fetal echocardiographic image segmentation using neural networks. 2[465], 507-511. 1999. Image Processing and its Applications 1999, Seventh International Conference.
4. Siqueira, M. L., Scharcanski, J., and Navaux, P. O. A., "Echocardiographic image sequence segmentation and analysis using self-organizing maps," *Journal of VLSI Signal Processing*, vol. 32 pp. 135-145, 2002.
5. Leventon, M. E., Grimson, W. E. L., and Faugeras, O. Statistical shape influence in geodesic active contours. 1, 316-321. 2000. Computer Vision and Pattern Recognition.
6. Dindoyal, I., Lambrou, T., Deng, J., Ruff, C. F., Linney, A. D., Rodeck, C. H., and Todd-Pokropek, A. Level set segmentation of the fetal heart. Frangi, A. F. 123-132. 2005. Barcelona, Spain, Springer-Verlag. Functional Imaging and Modelling of the Heart. 2-6-2005.
7. Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W. E., and Willsky, A., "A shape-based approach to the segmentation of medical imagery using level sets," *IEEE Transactions on Medical Imaging*, vol. 22, no. 2, pp. 137-154, 2003.
8. Storn, R. and Price, K. Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces. 1995.